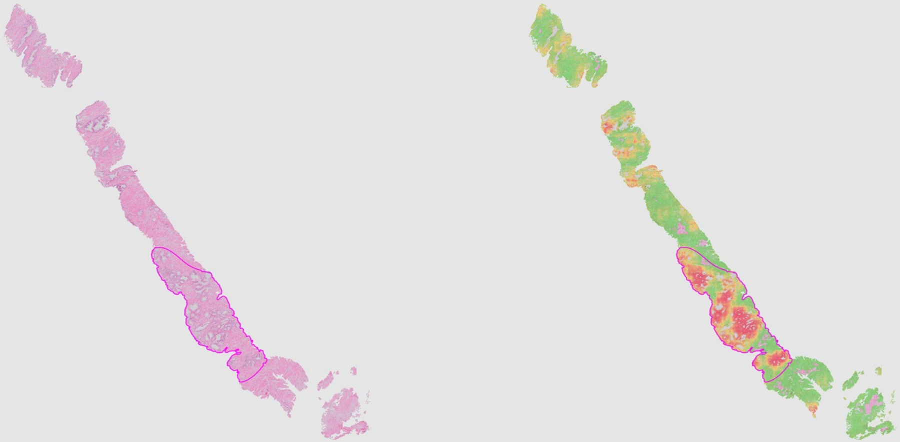
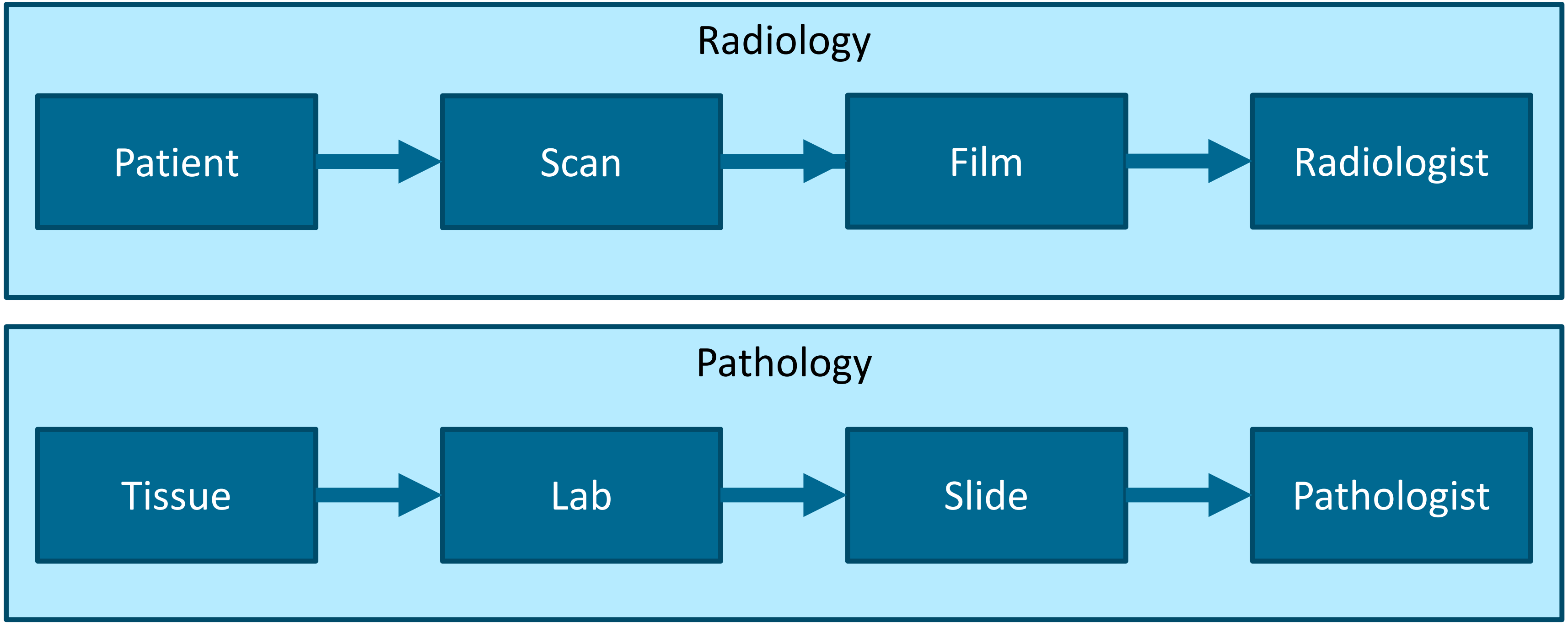
Applications of Machine Learning for Clinical Practice



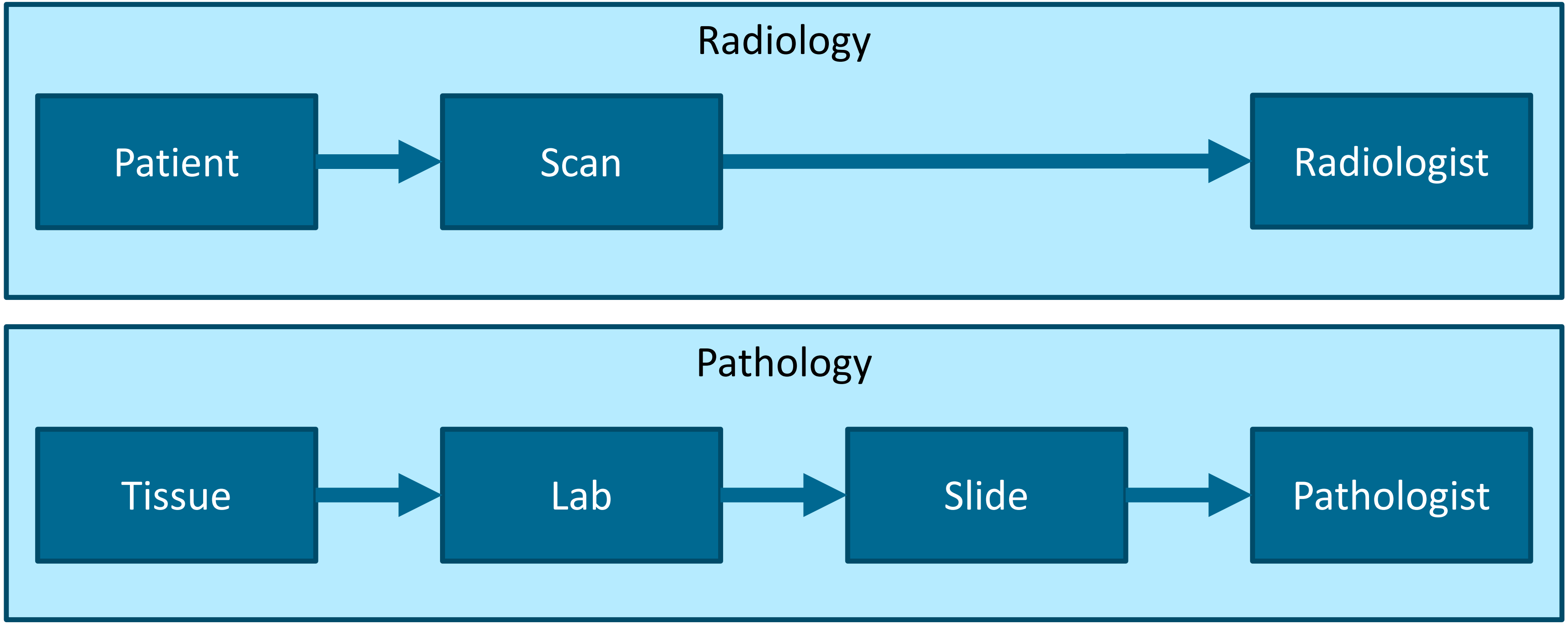
Geert Litjens
Department of Pathology
Radboud University Medical Center

The promise of digital pathology?

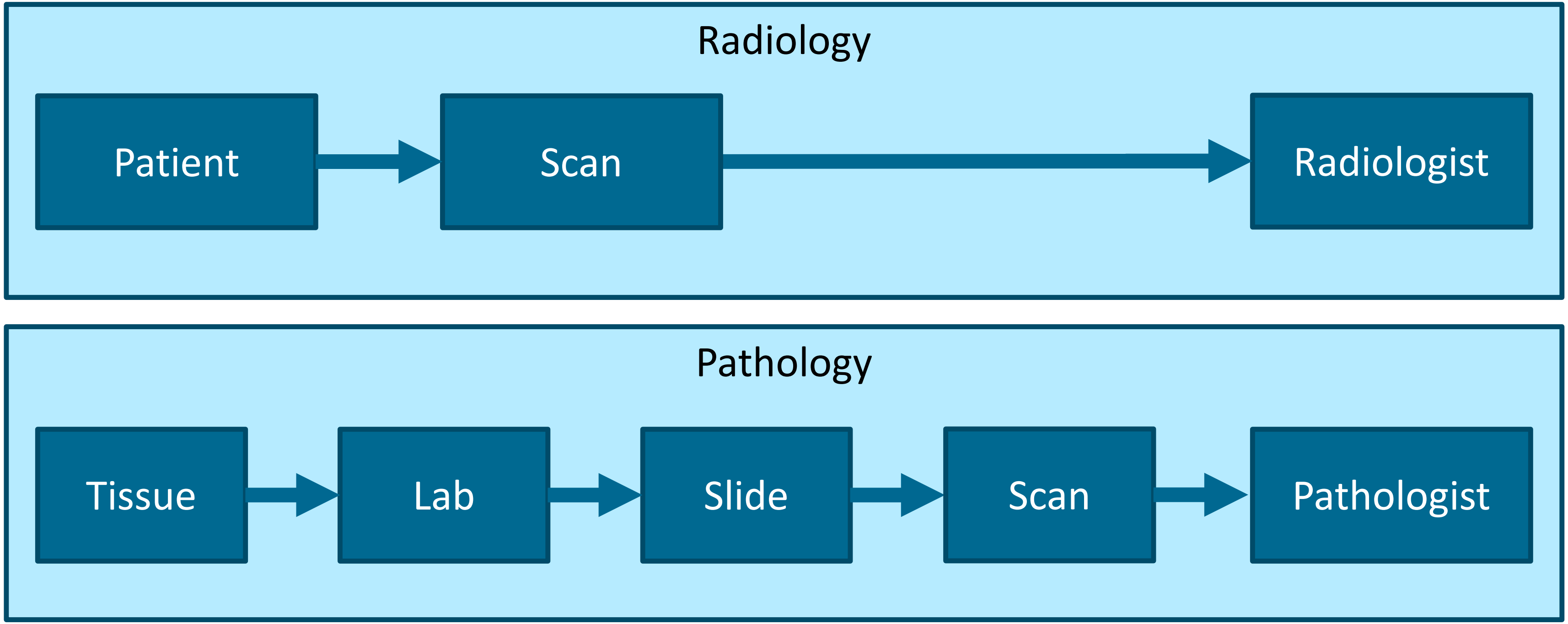
The promise of digital pathology?



The promise of digital pathology?



The promise of digital pathology?



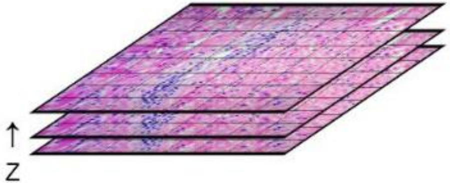
The promise of digital pathology?

Scanners

Storage

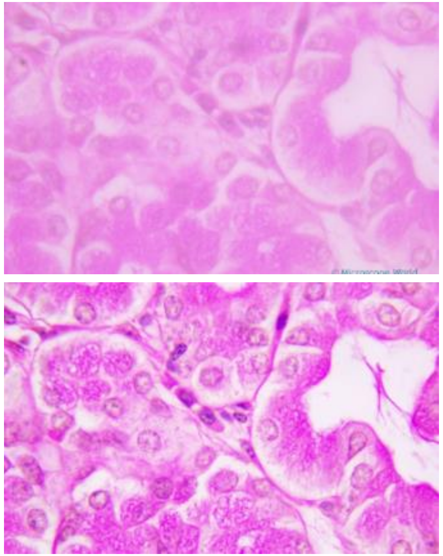
Computers

The promise of digital pathology?



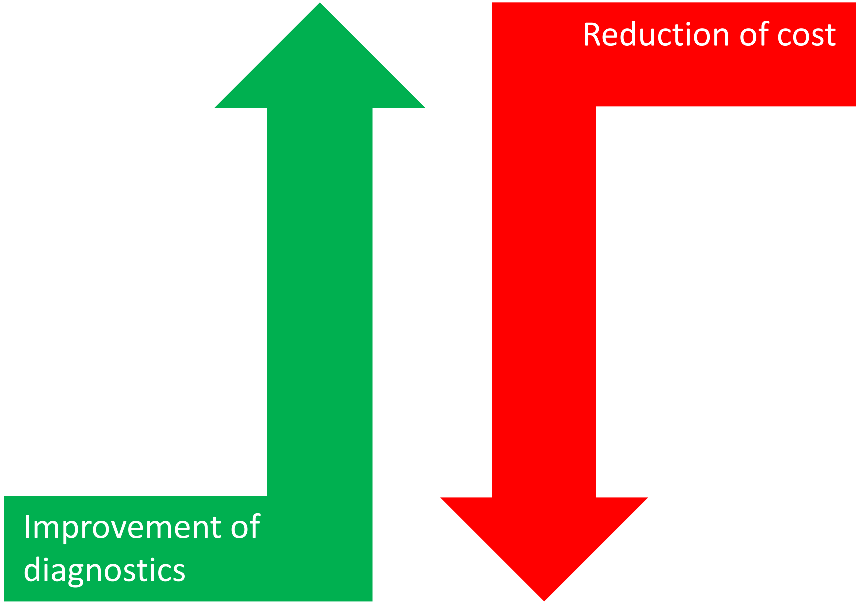
Multiple focal points

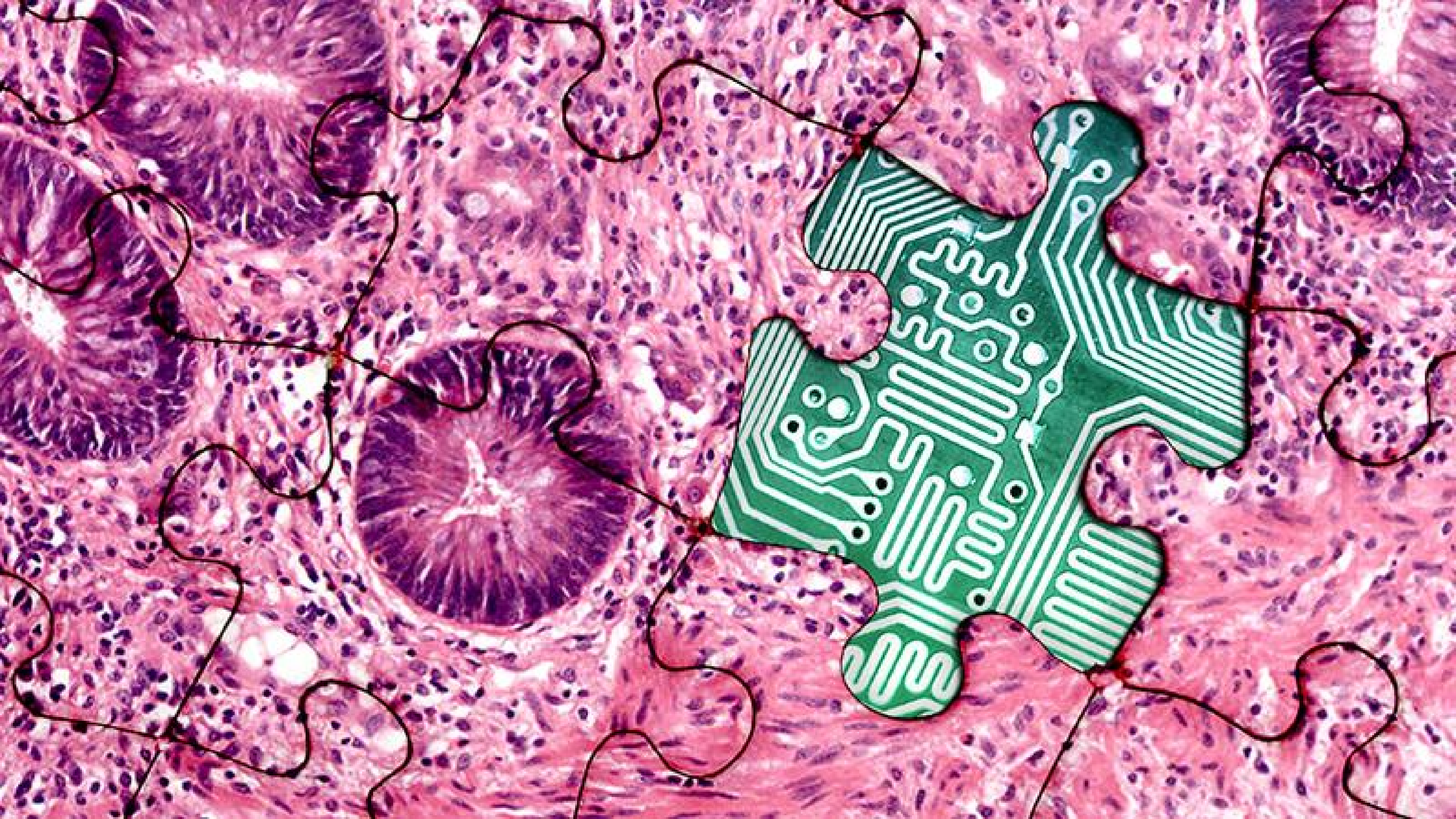
Large slides



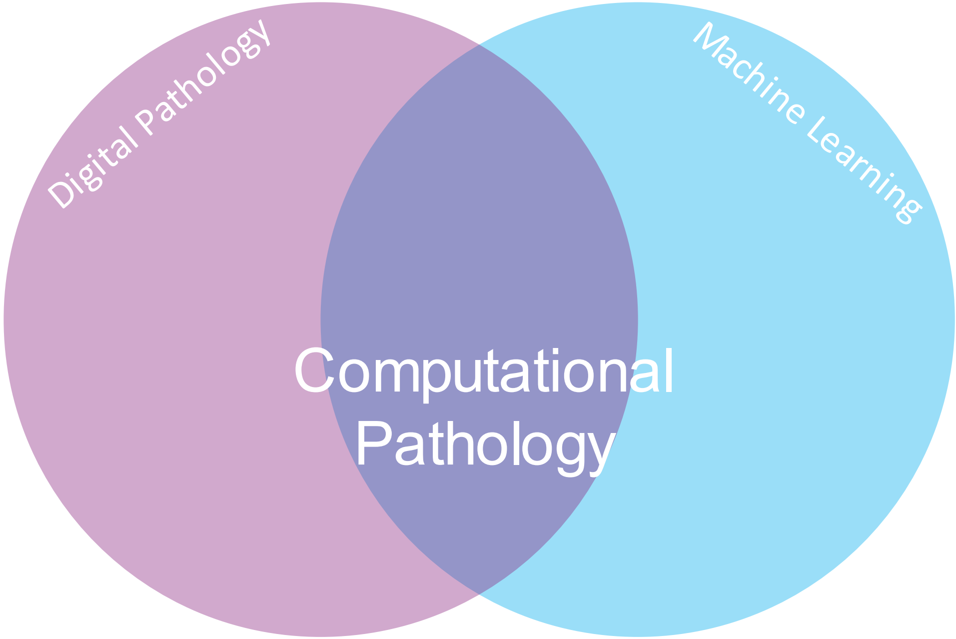
Oil immersion

The promise of digital pathology?

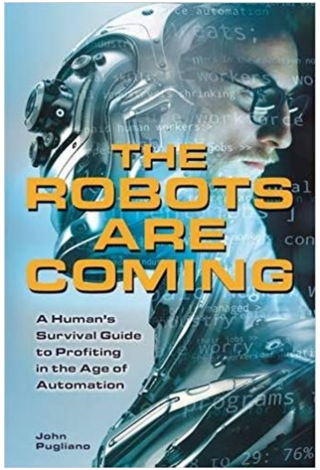




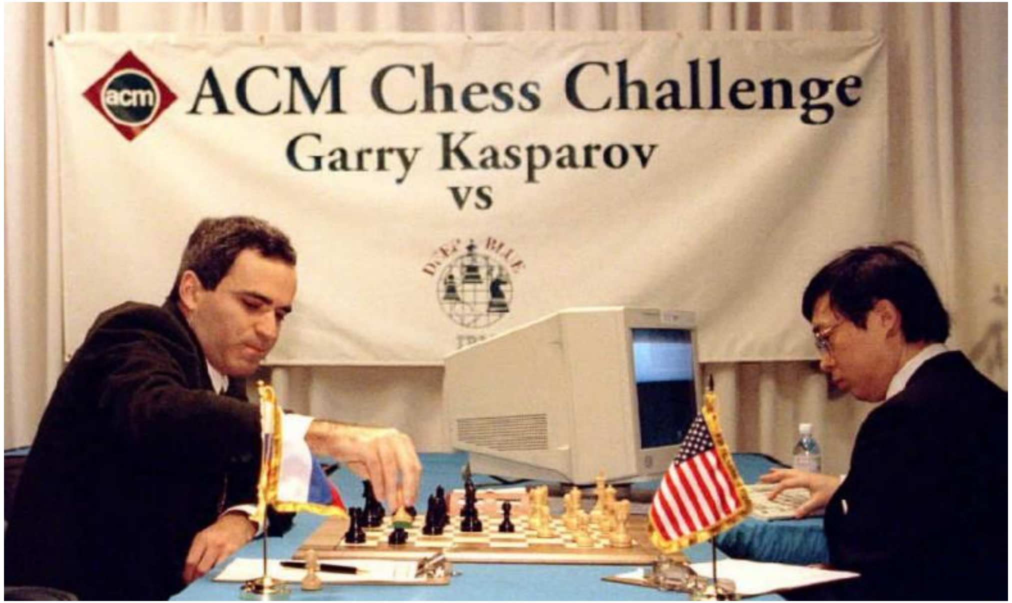
Computational Pathology



Machine learning



ML: a bit of history



nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

At last — a computer program that can beat a champion Go player PAGE 484

ALL SYSTEMS GO

CONSERVATION
SONGBIRDS À LA CARTE
Illegal harvest of millions of Mediterranean birds
PAGE 452

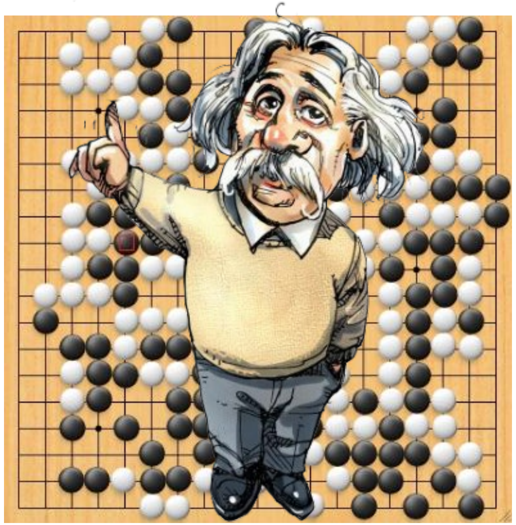
RESEARCH ETHICS
SAFEGUARD TRANSPARENCY
Don't let openness backfire on individuals
PAGE 459

POPULAR SCIENCE
WHEN GENES GOT 'SELFISH'
Dawkins's calling card forty years on
PAGE 462

NATURE.COM/NATURE
28 January 2016 £10
Vol. 529, No. 7587

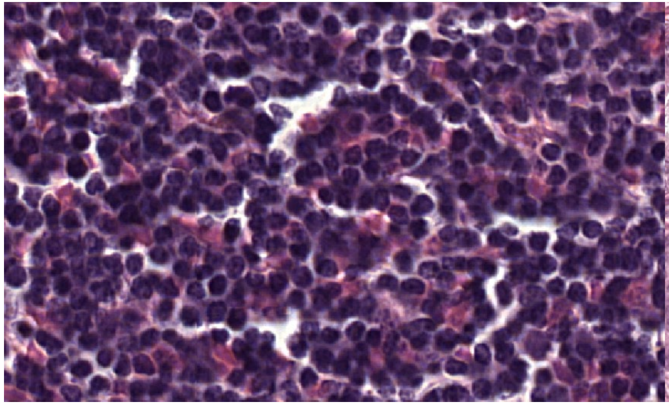
ML: a bit of history

30 possible moves per turn
40 turns per game

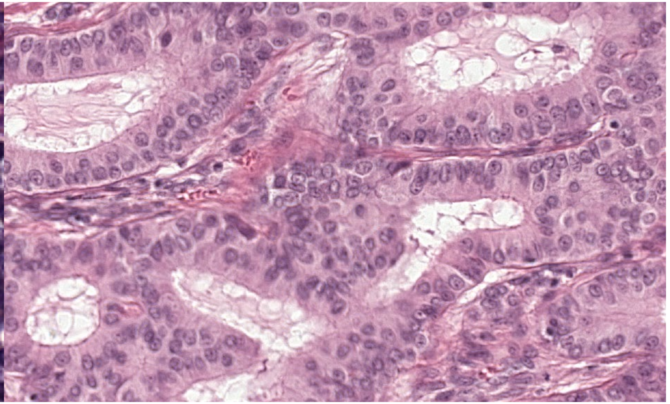


250 possible moves per turn
150 turns per game

How to build an ML system?



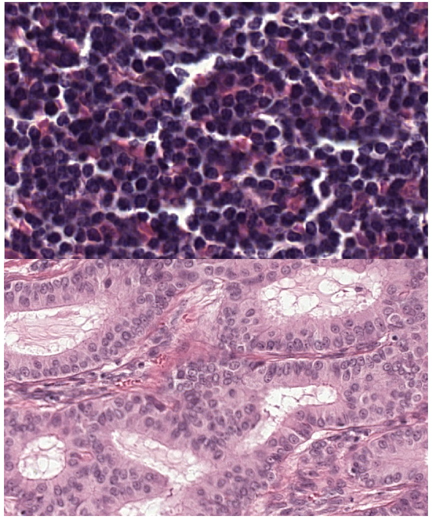
Normal lymph node



Breast cancer metastasis

How to build an ML system?

Examples

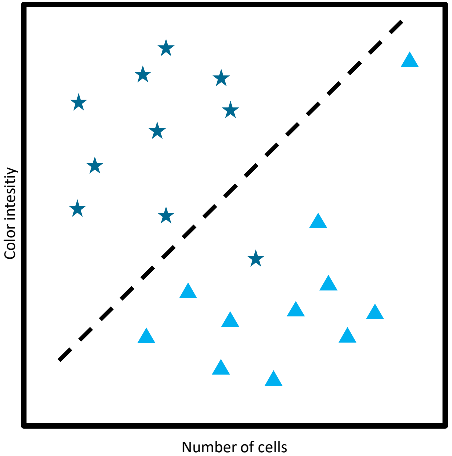


Features

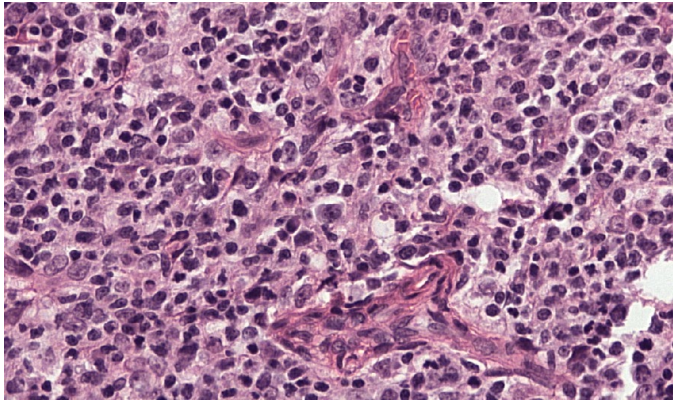
Color intesitiy

Number of cells

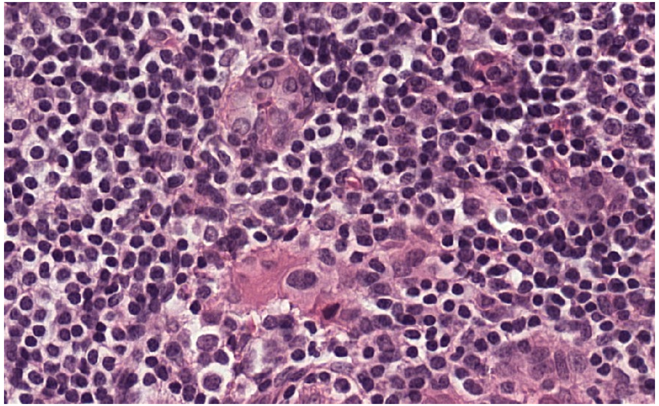
Classification



How to build an ML system?



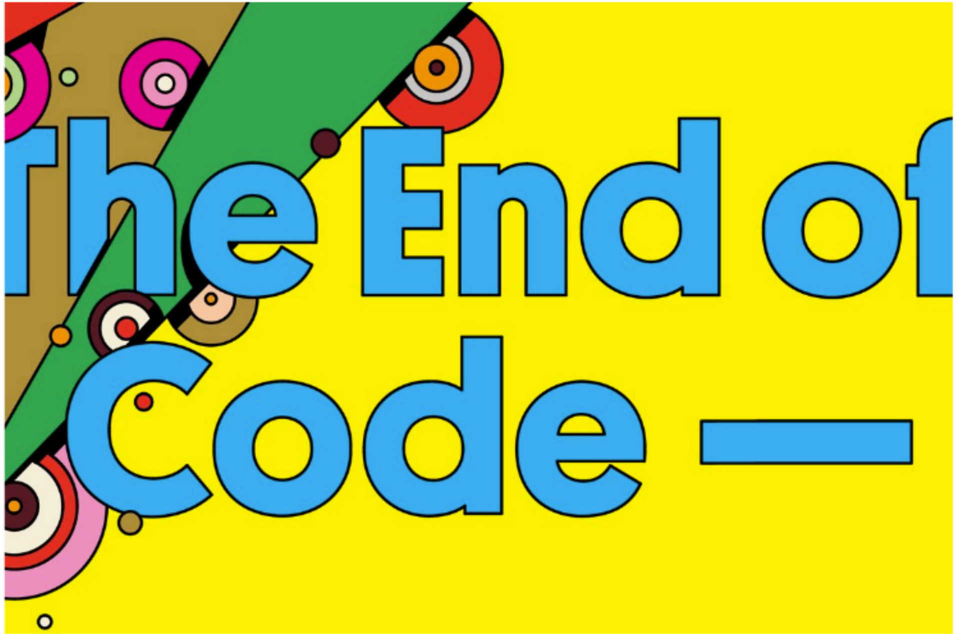
Normal lymph node



Breast cancer metastasis

JASON TANZ IDEAS 05.17.16 06:50 AM

SOON WE WON'T PROGRAM COMPUTERS. WE'LL TRAIN THEM LIKE DOGS



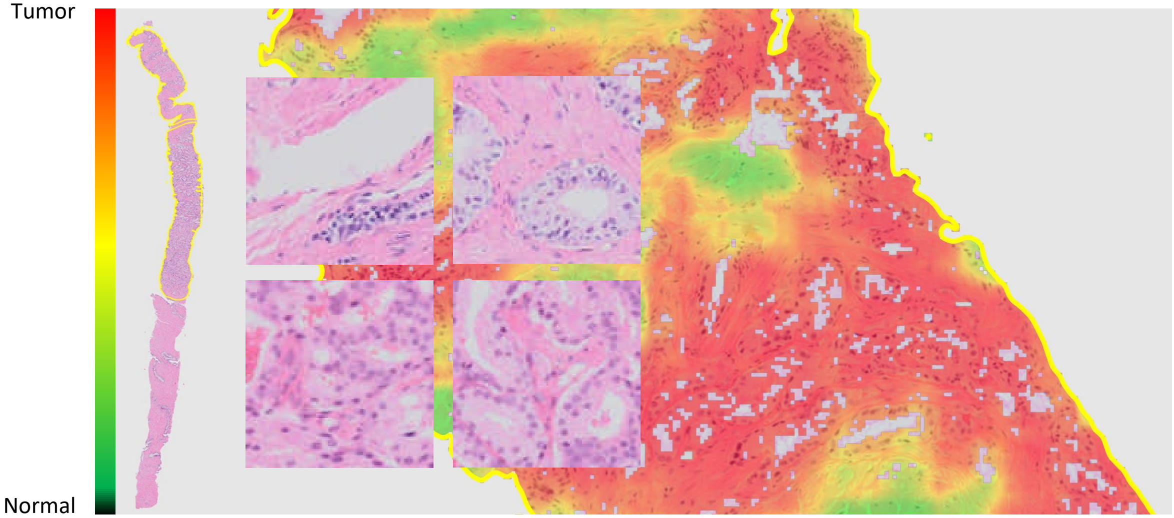
EDWARD C. MONAGHAN

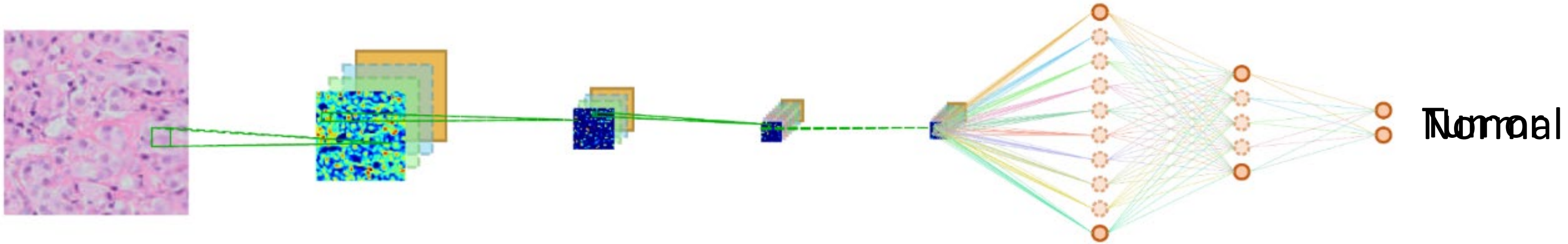
BEFORE THE INVENTION of the computer, most experimental psychologists thought the brain was an unknowable black box. You could analyze a subject's behavior—*ring bell, dog salivates*—but thoughts, memories, emotions? That stuff was obscure and inscrutable, beyond the reach of science. So these behaviorists, as they called themselves, confined their work to the study of stimulus and response, feedback and reinforcement, bells and saliva. They gave up trying to

JASON TANZ IDEAS 05.17.16 06:50 AM

SOON WE WON'T PROGRAM COMPUTERS. WE'LL TRAIN THEM LIKE DOGS

How to build an ML system?


Tumor
Normal


Tumor
Normal

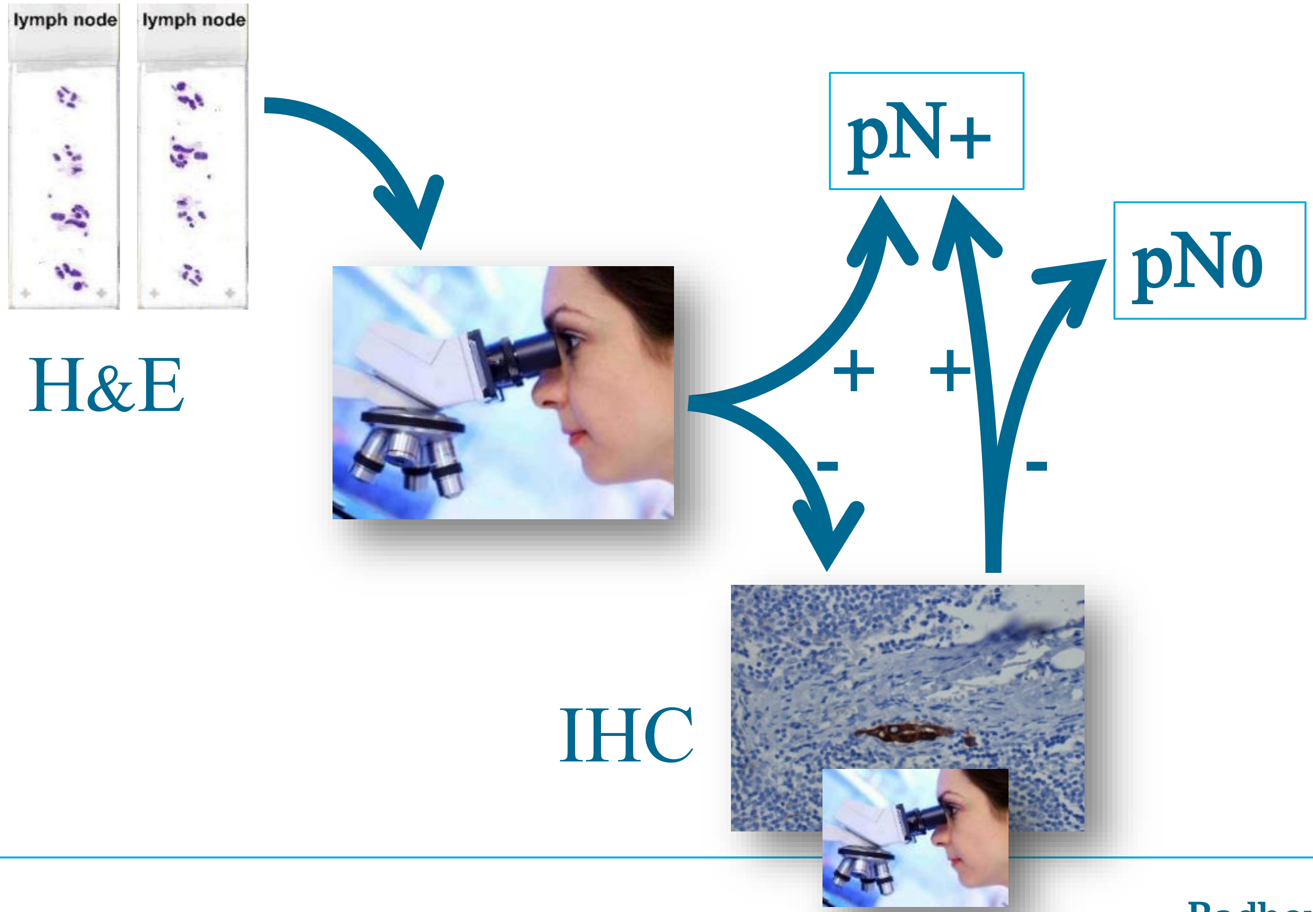
Practical applications of computation pathology

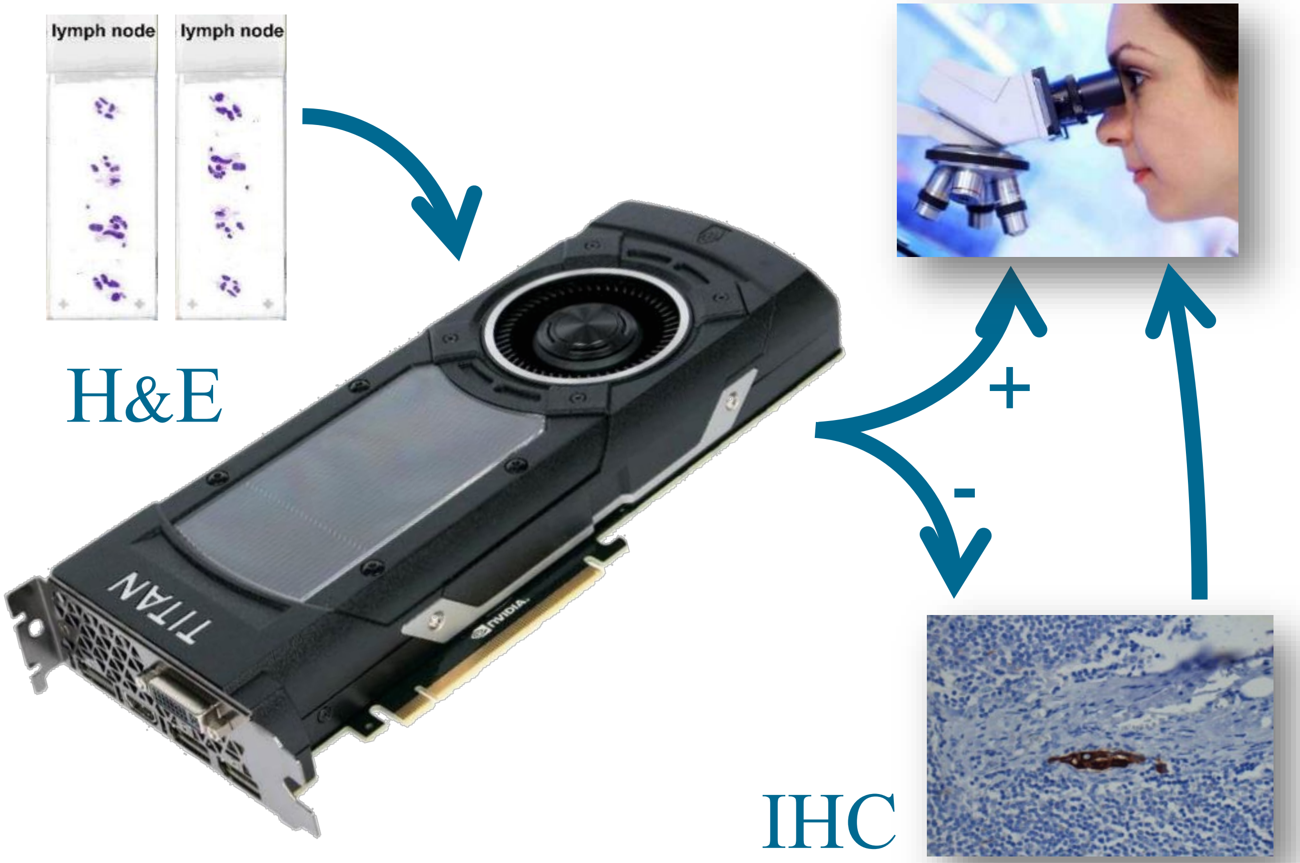
Detection of metastases in lymph nodes

Automatic mitotic counts

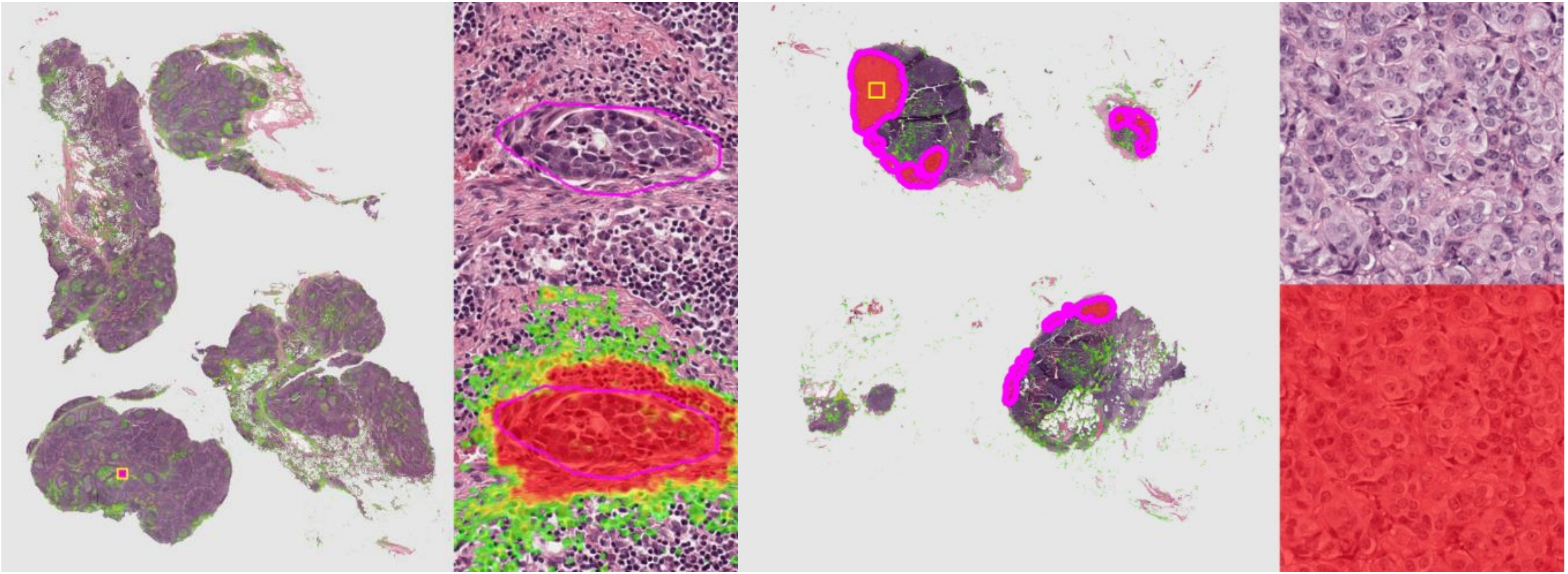
Tumor
qua

Detection of metastases in lymph nodes


lymph node
lymph node
H&E
pN+
pN0
+
+
-
-
IHC


lymph node
lymph node
H&E
TITAN
NVIDIA
+
-
IHC

Detection of metastases in lymph nodes



Breast cancer metastasis detection

Data

Centrum	Number of slides
CWZ (Nijmegen)	200
LabPON (Hengelo)	200
Rijnstate (Arnhem)	200
Radboudumc (Nijmegen)	439
UMCU (Utrecht)	350
Total	**1399**


CAMELYON16


CAMELYON17

Why challenges?

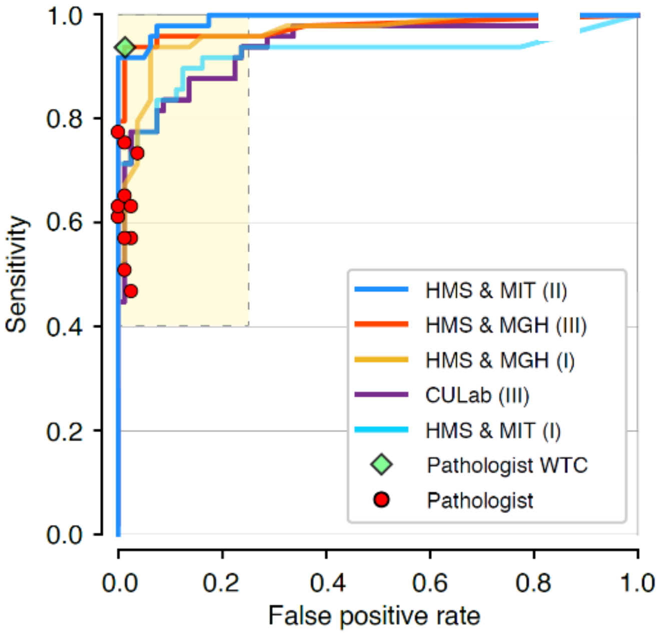
Great way to collect and compare solutions for a problem

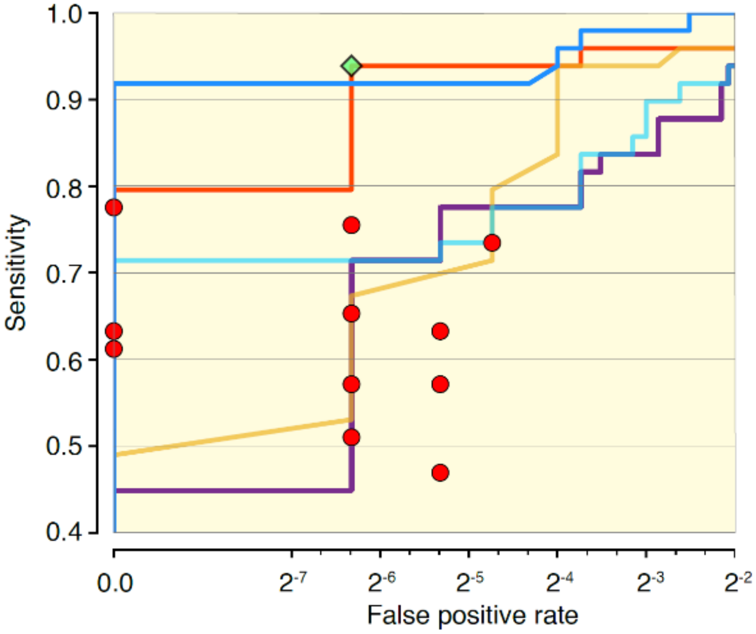
Fair comparison of algorithms

- Same evaluation metric
- Same ground truth definition
- Same training and test datasets

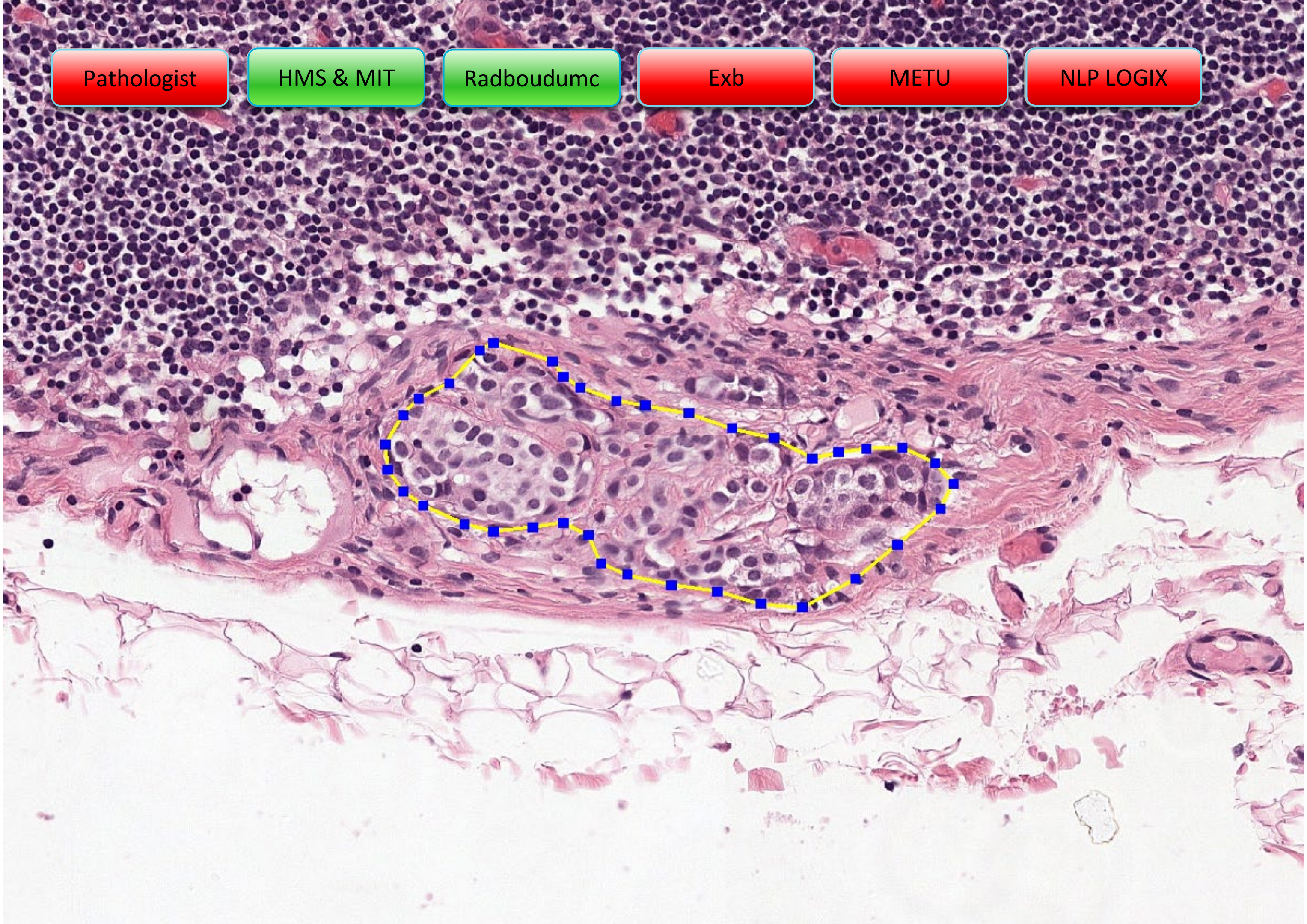
Rank	Team	AUC	Description
01	**Harvard Medical School (BIDMC) and Massachusetts Institute of Technology (CSAIL), USA**	**0.9250**	
02	ExB Research and Development co., Germany	0.9173	
03	Independent participant, Germany	0.8680	
04	Health Sciences Middle East Technical University, Turkey	0.8669	
05	NLP LOGIX co., USA	0.8332	
06	University of Toronto, Electrical and Computer Engineering, Canada	0.8181	
07	The Warwick-QU Team, United Kingdom	0.7999	
08	Radboud University Medical Center, Diagnostic Image Analysis Group, Netherlands	0.7828	
09	HTW-BERLIN, Germany	0.7717	
10	University of Toronto, Electrical and Computer Engineering, Canada	0.7666	

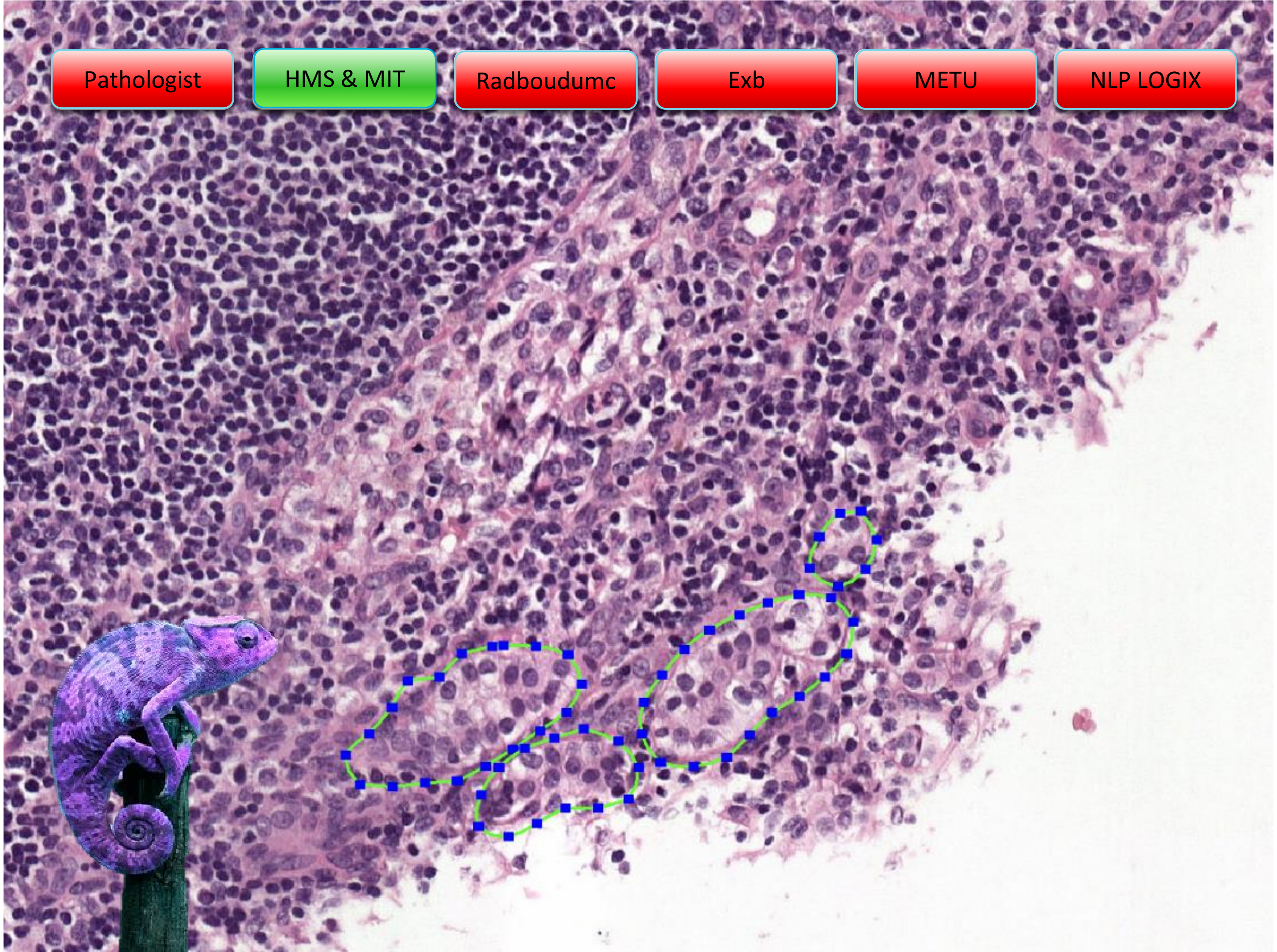
Comparing to pathologists

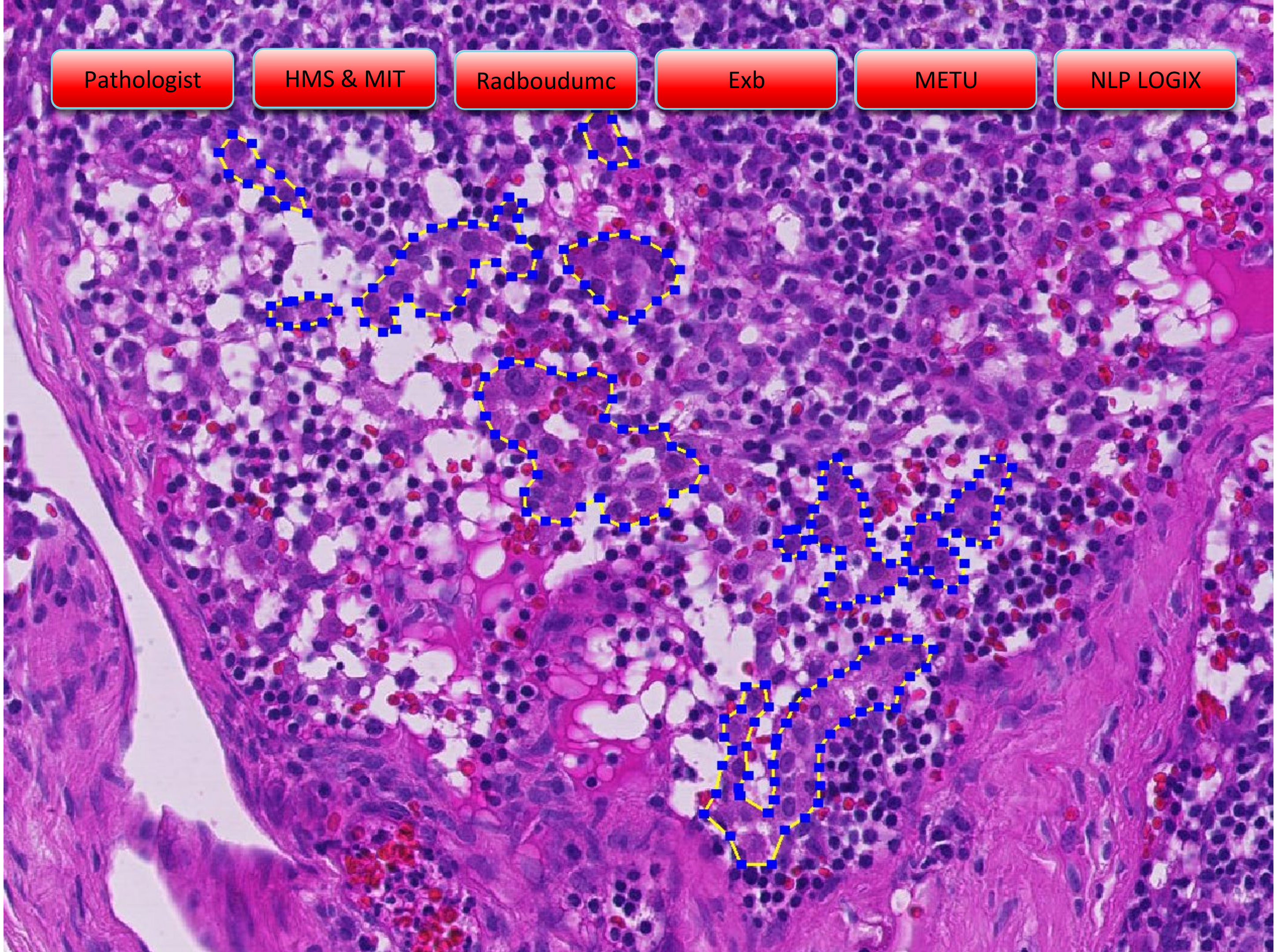


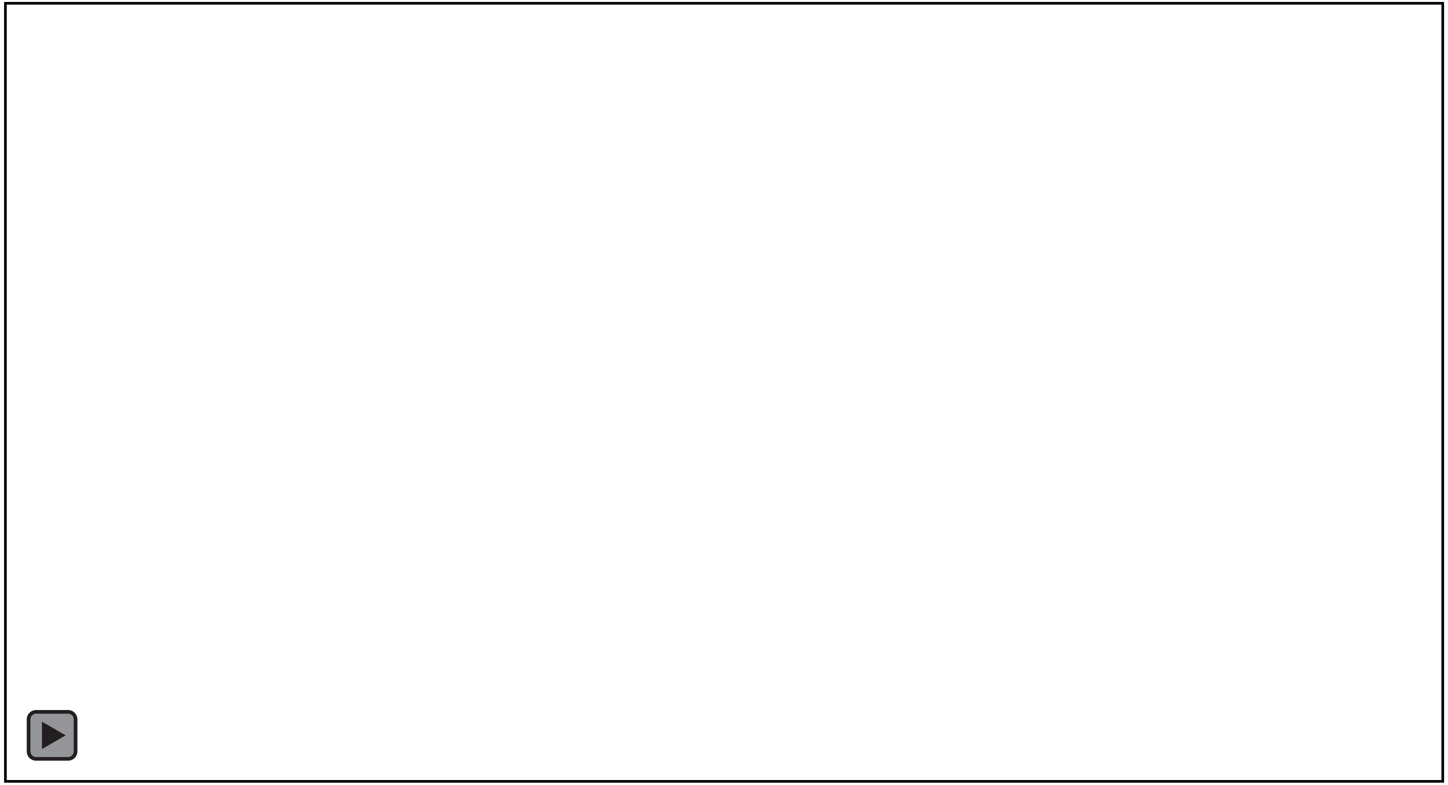


Ehteshami et al. *JAMA. 2017*


Pathologist
HMS & MIT
Radboudumc
Exb
METU
NLP LOGIX


Pathologist
HMS & MIT
Radboudumc
Exb
METU
NLP LOGIX


Pathologist
HMS & MIT
Radboudumc
Exb
METU
NLP LOGIX



Implemented in clinical practice

62 patients

31 patients

16 N0	1 N0(i+)	3 N1mi
4 N1	7 errors	

1 micrometastasis missed
1 micrometastasis misclassified as ITC
5 ITCs missed (only visible on IHC in retrospect)

Microscope

31 patients

14 N0	0 N0(i+)	4 N1mi
4 N1	9 errors	

65% less time needed

Bult et al. *In preparation*

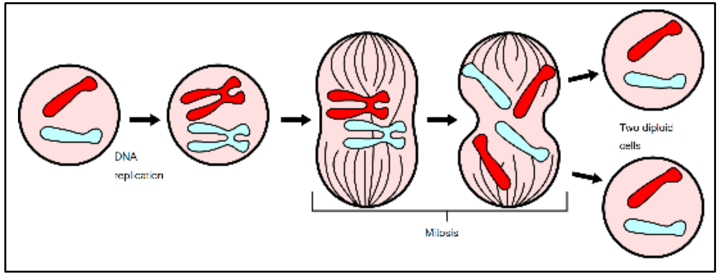
Practical applications of computation pathology

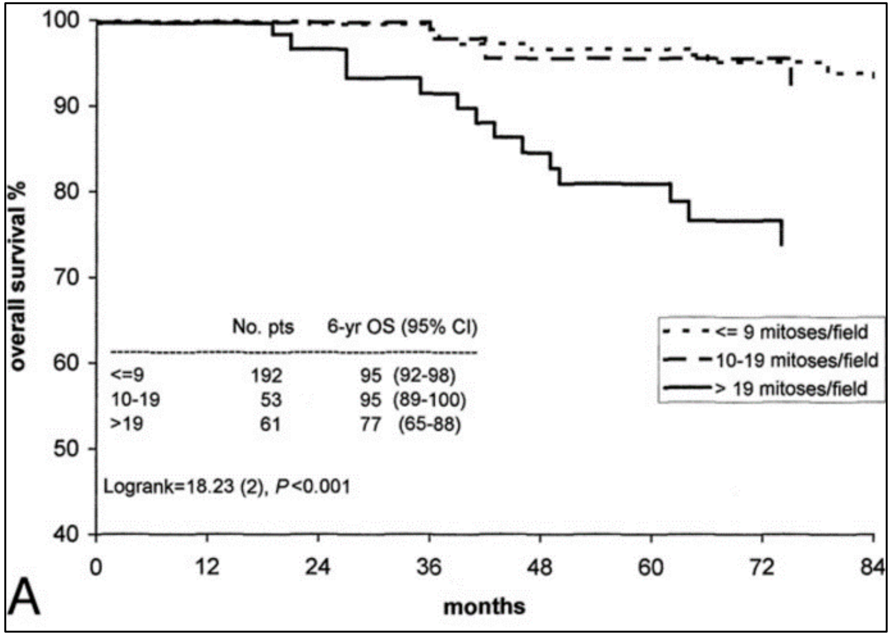
Detection of metastases in lymph nodes

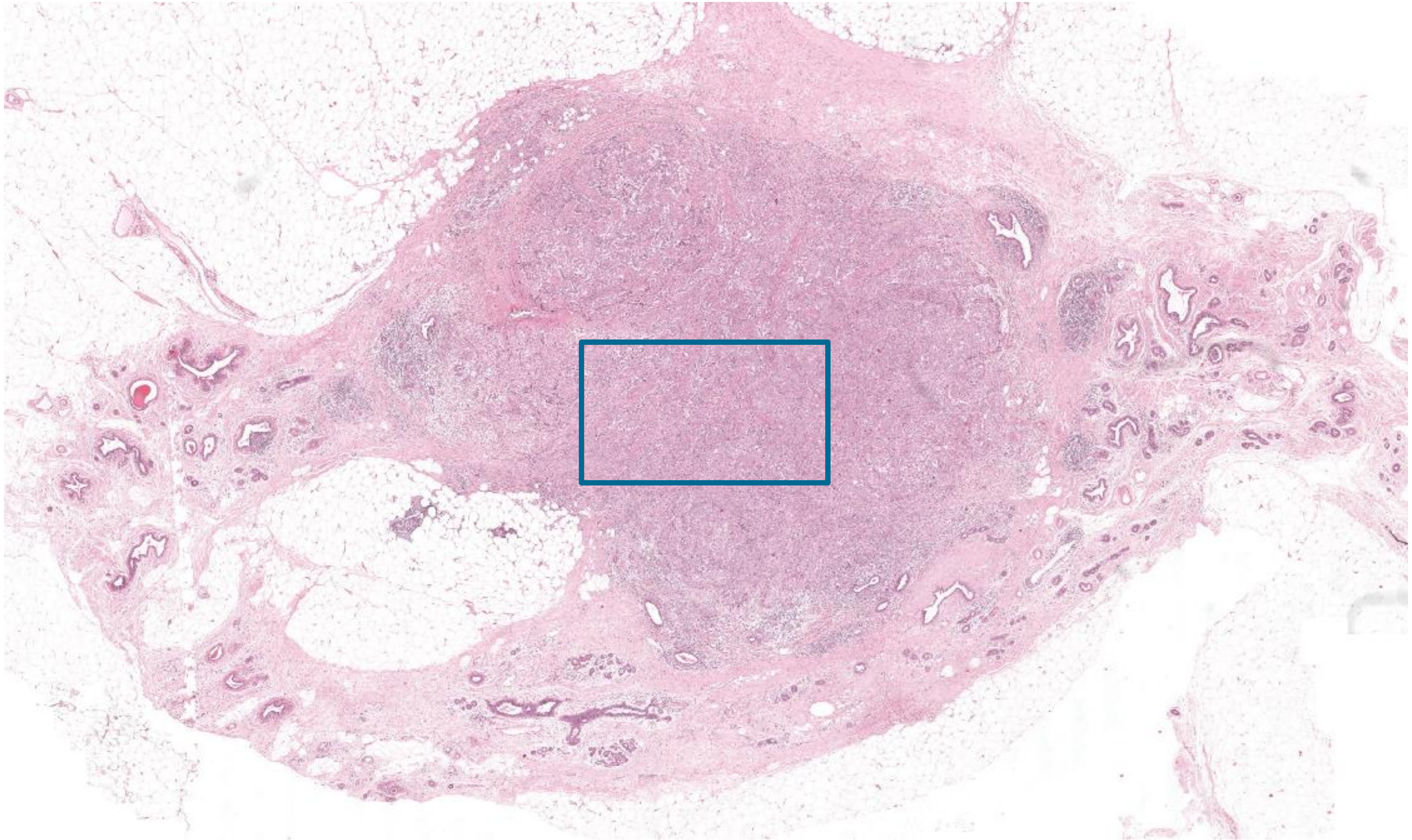
Automatic mitotic counts

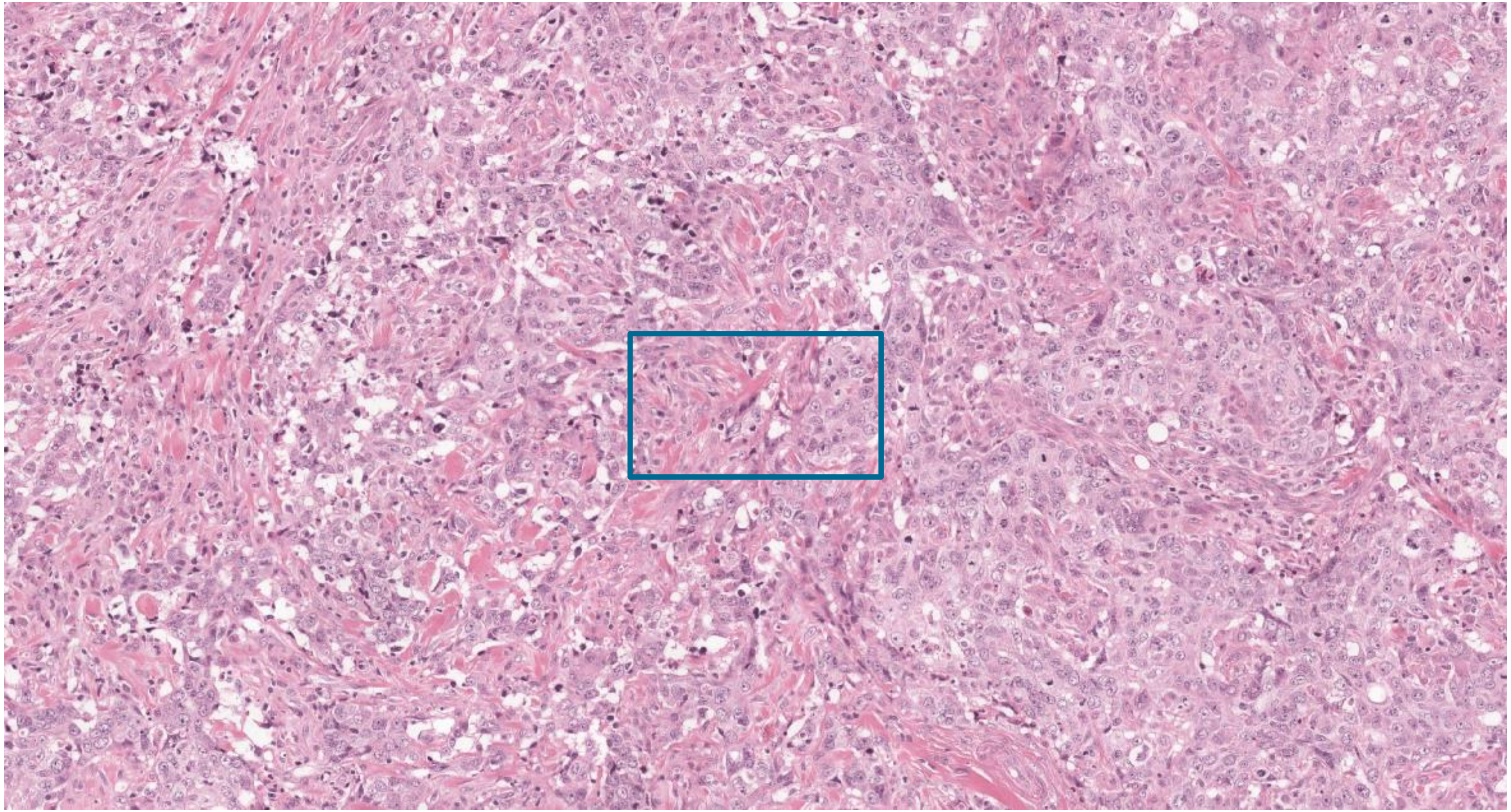
Tumor
qua

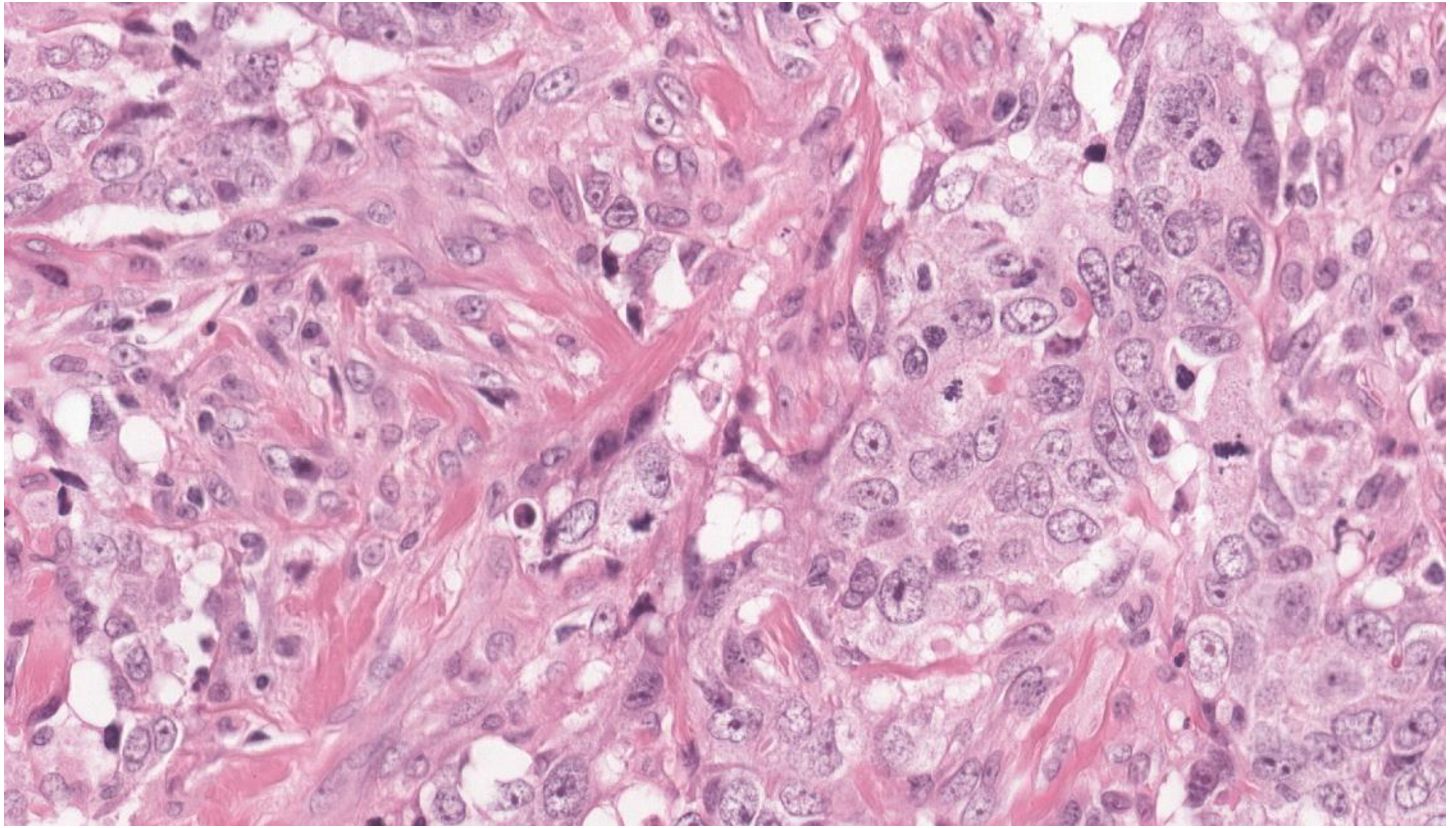
Automatic mitotic counts











Automatic mitotic counts

Assessment of Mitosis Detection Algorithms 2013

AMIDA13 | MICCAI Grand Challenge

D. C. Cireşan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Mitosis detection in breast cancer histology images with deep neural networks," in *International Conference on Medical Image Computing and Computer-assisted Intervention.* Springer, 2013, pp. 411–418.

M. Veta, P. J. van Diest, M. Jiwa, S. Al-Janabi, and J. P. Pluim, "Mitosis counting in breast cancer: Object-level interobserver agreement and comparison to an automatic method," *PloS one*, vol. 11, no. 8, p. e0161286, 2016.

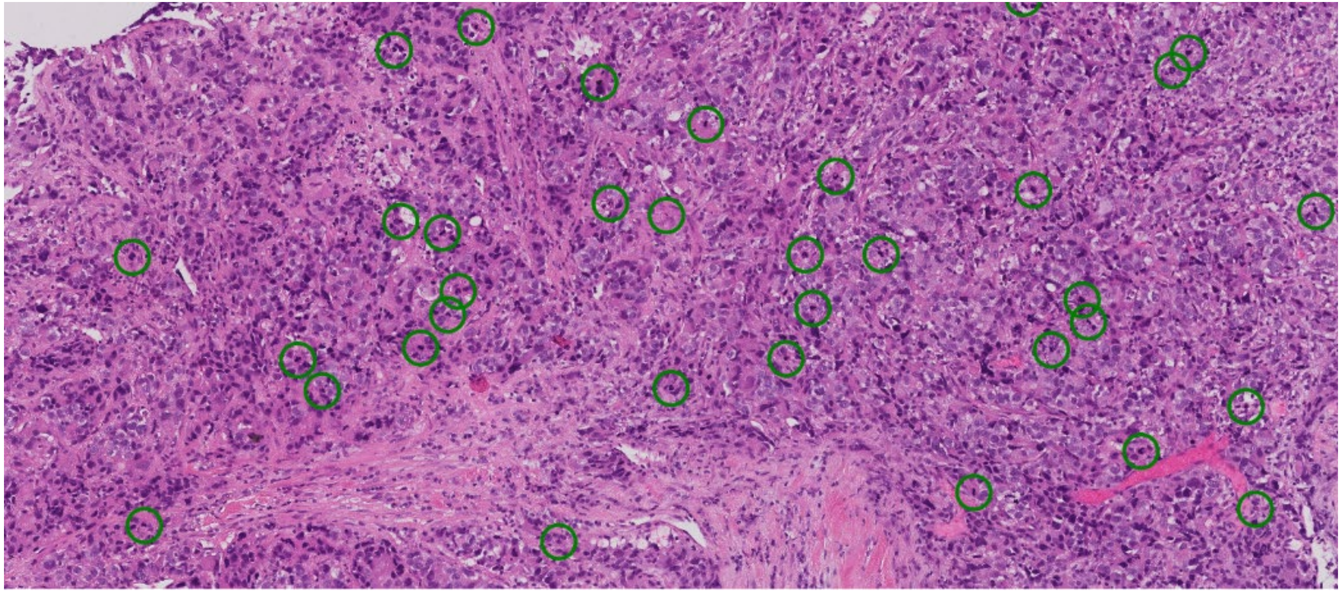
Tumor Proliferation Assessment Challenge 2016

TUPAC16 | MICCAI Grand Challenge

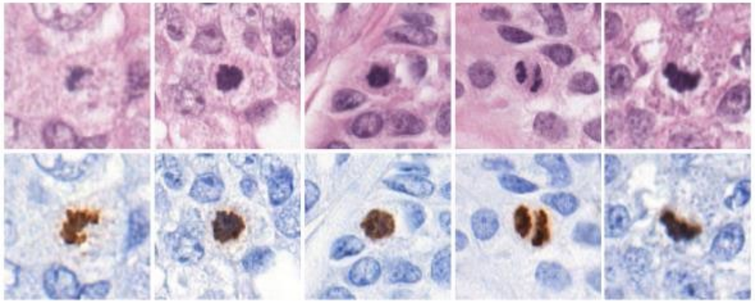
E. Zerhouni, D. Lányi, M. Viana, and M. Gabrani, "Wide residual networks for mitosis detection," in *Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on.* IEEE, 2017, pp. 924–928.

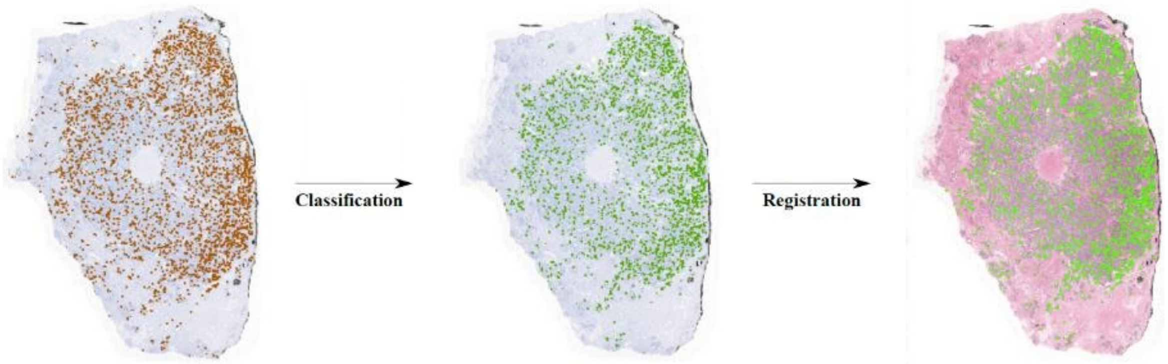
K. Paeng, S. Hwang, S. Park, M. Kim, and S. Kim, "A unified framework for tumor proliferation score prediction in breast histopathology," *arXiv preprint arXiv:1612.07180*, 2016.

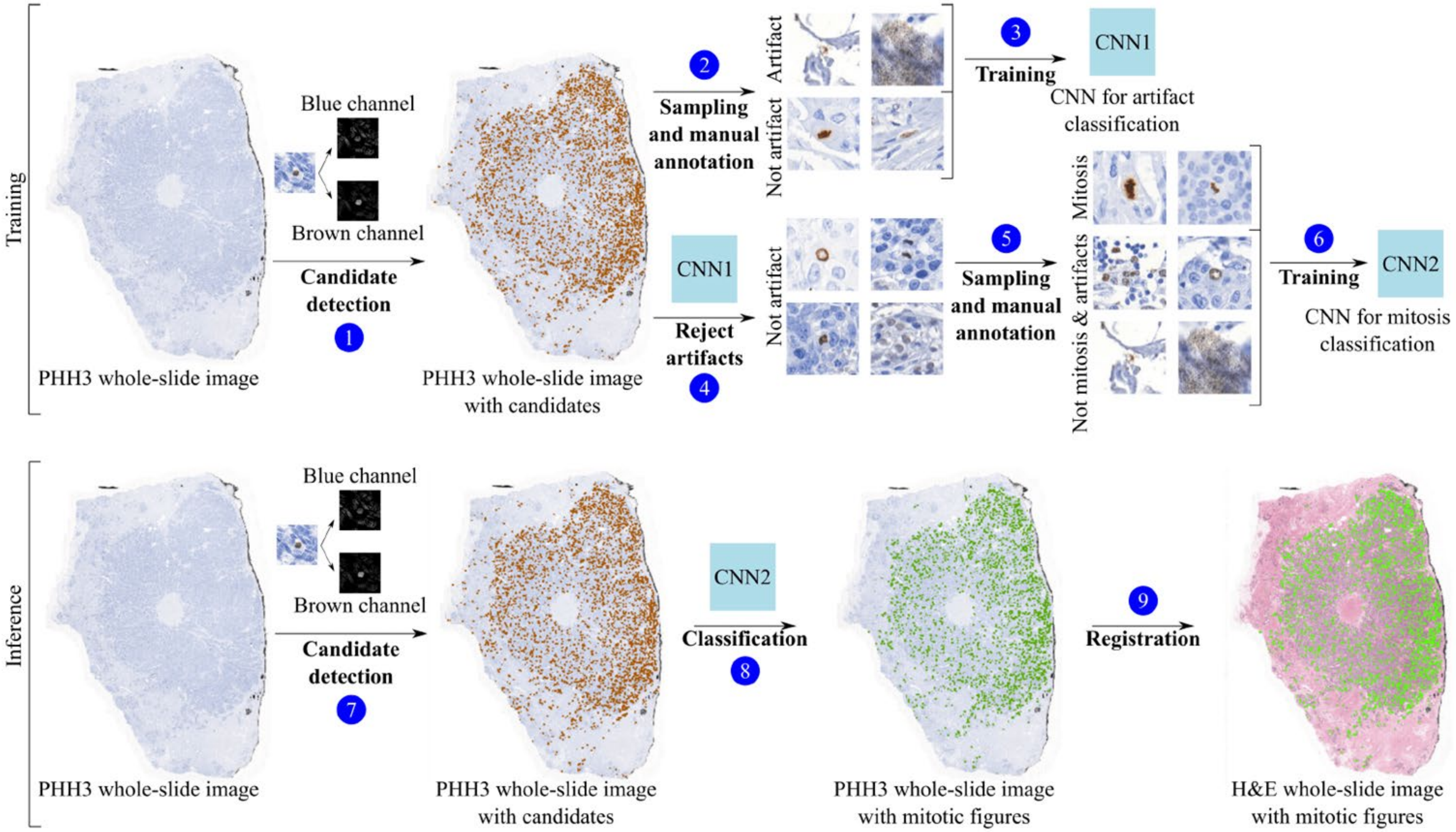
Challenge 1: Reference standard

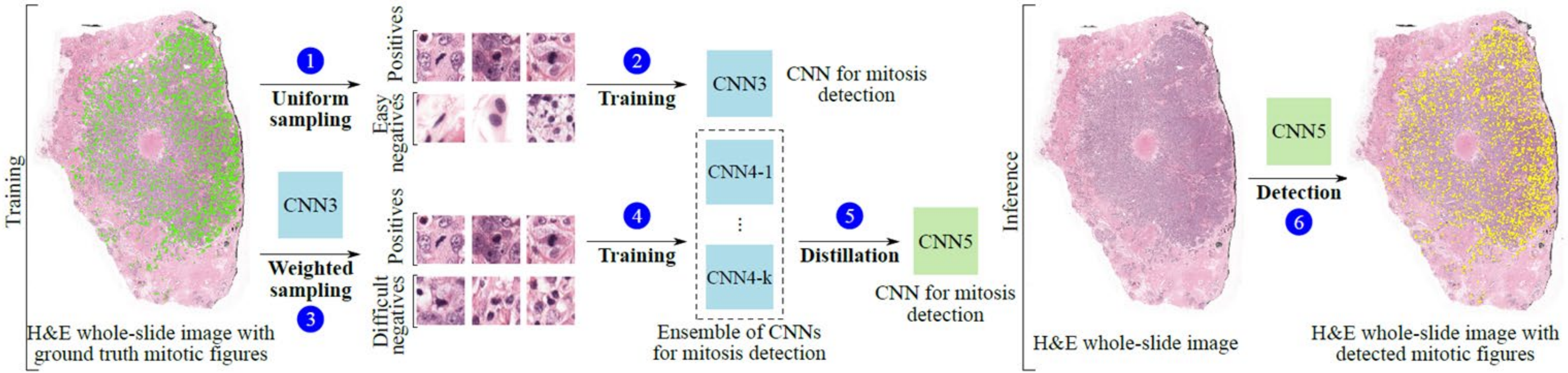


IHC offers a solution



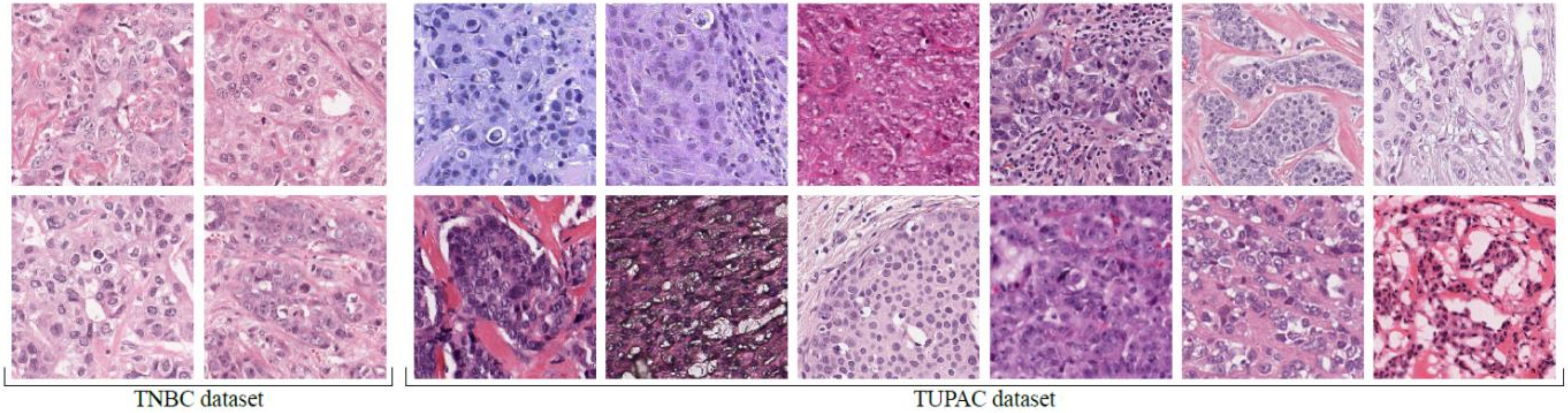



Training
Blue channel
Brown channel
Candidate detection
1
PHH3 whole-slide image
PHH3 whole-slide image with candidates
2
Sampling and manual annotation
Artifact
Not artifact
3
Training
CNN1
CNN for artifact classification
CNN1
Reject artifacts
4
Not artifact
5
Sampling and manual annotation
Mitosis
Not mitosis & artifacts
6
Training
CNN2
CNN for mitosis classification
Inference
Blue channel
Brown channel
Candidate detection
7
PHH3 whole-slide image
PHH3 whole-slide image with candidates
CNN2
Classification
8
PHH3 whole-slide image with mitotic figures
9
Registration
H&E whole-slide image with mitotic figures

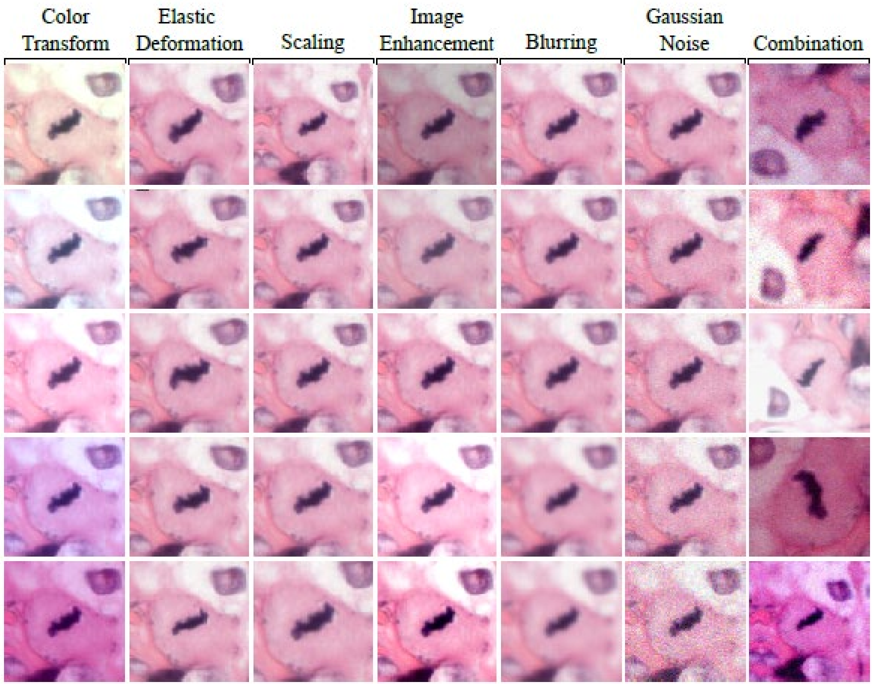


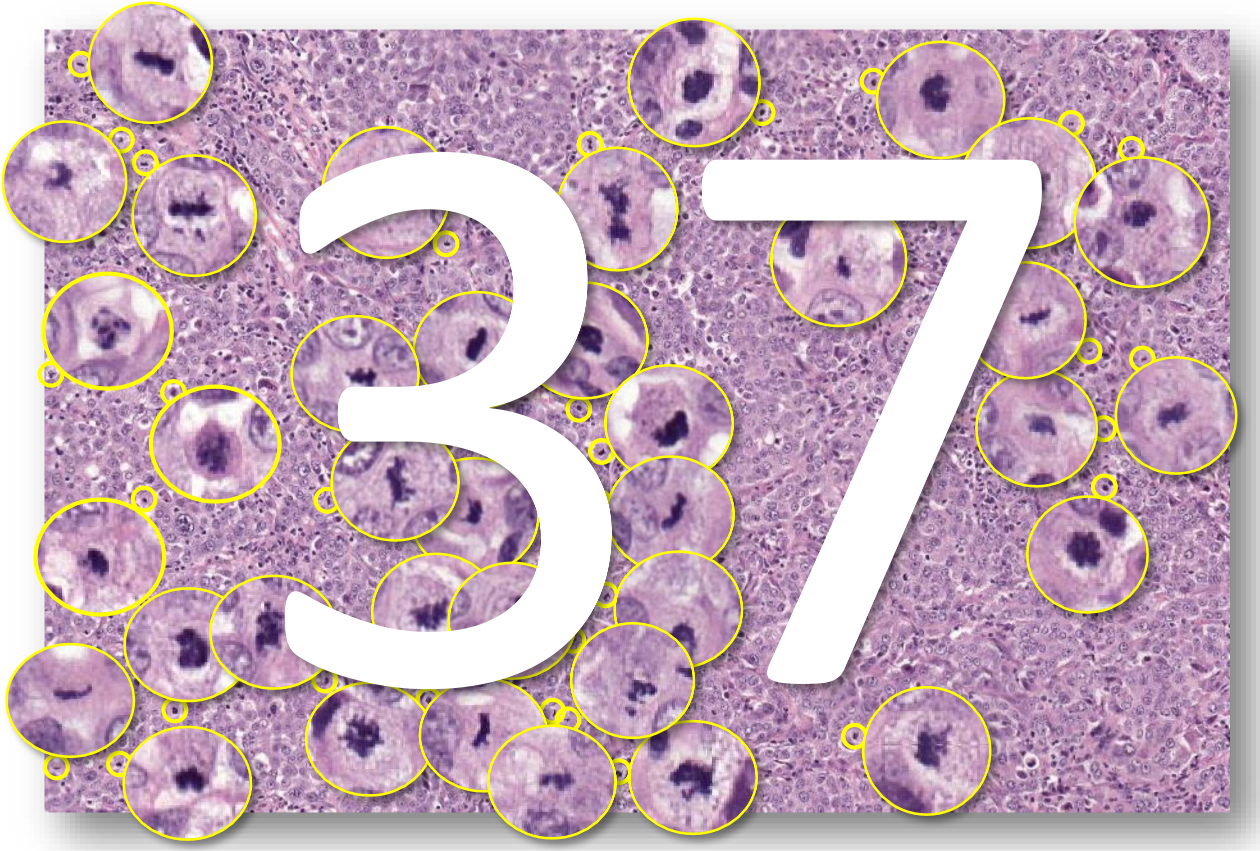
Training
H&E whole-slide image with ground truth mitotic figures
1
Uniform sampling
CNN3
Weighted sampling
3
Positives
Easy negatives
Positives
Difficult negatives
2
Training
4
Training
CNN3
CNN for mitosis detection
CNN4-1
CNN4-k
Ensemble of CNNs for mitosis detection
5
Distillation
CNN5
CNN for mitosis detection
Inference
H&E whole-slide image
CNN5
Detection
6
H&E whole-slide image with detected mitotic figures

Challenge 2: staining differences

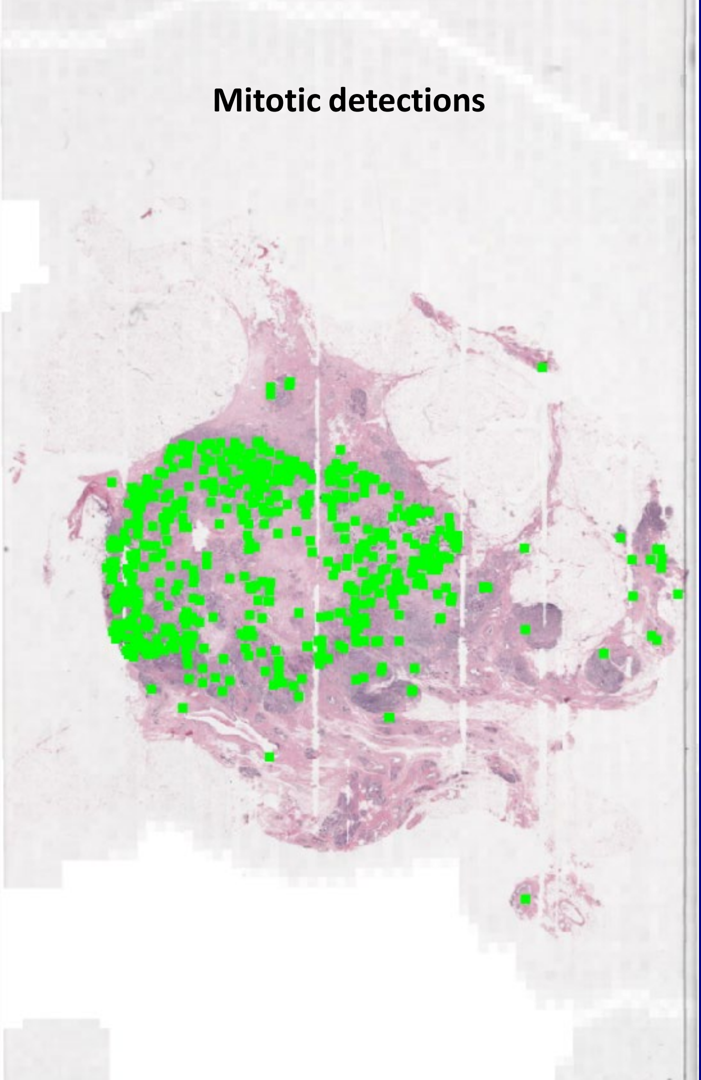


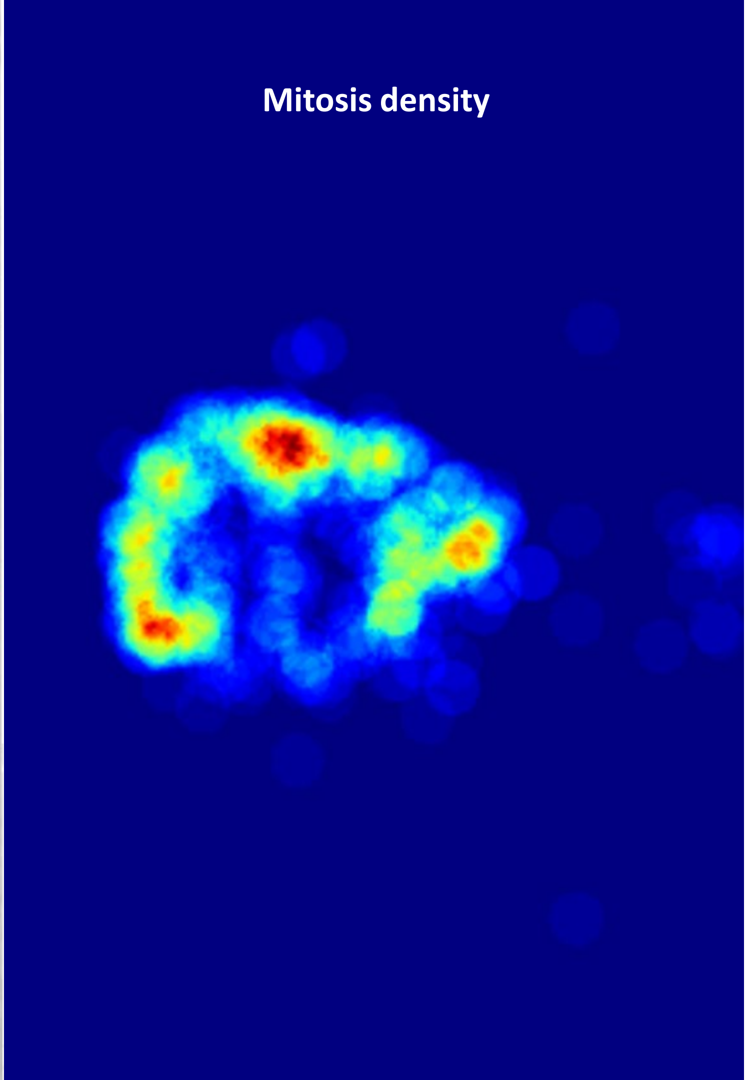
Solution 2: Data augmentation

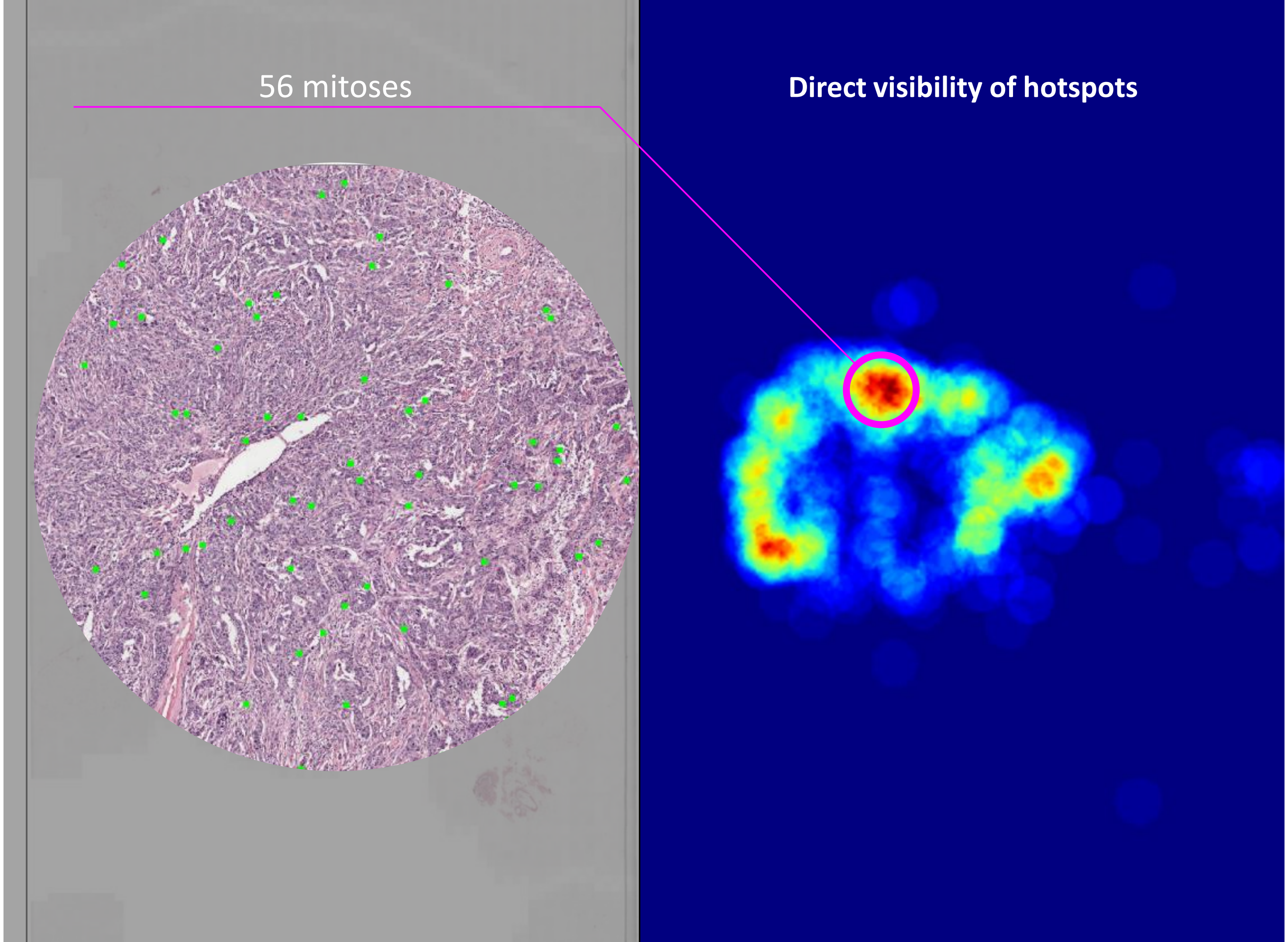


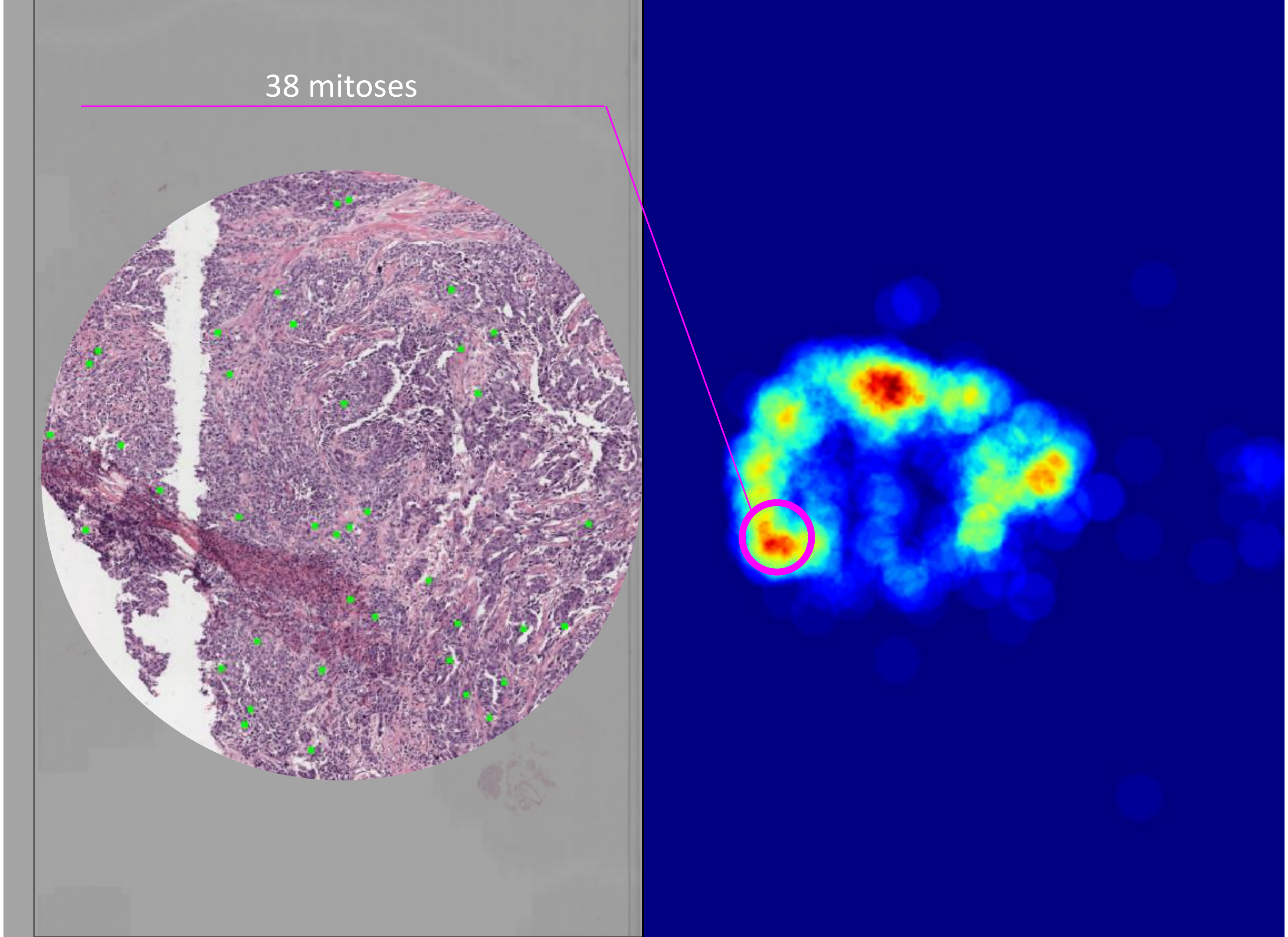


Tellez et al. *IEEE Transactions on Medical Imaging.* (2018)

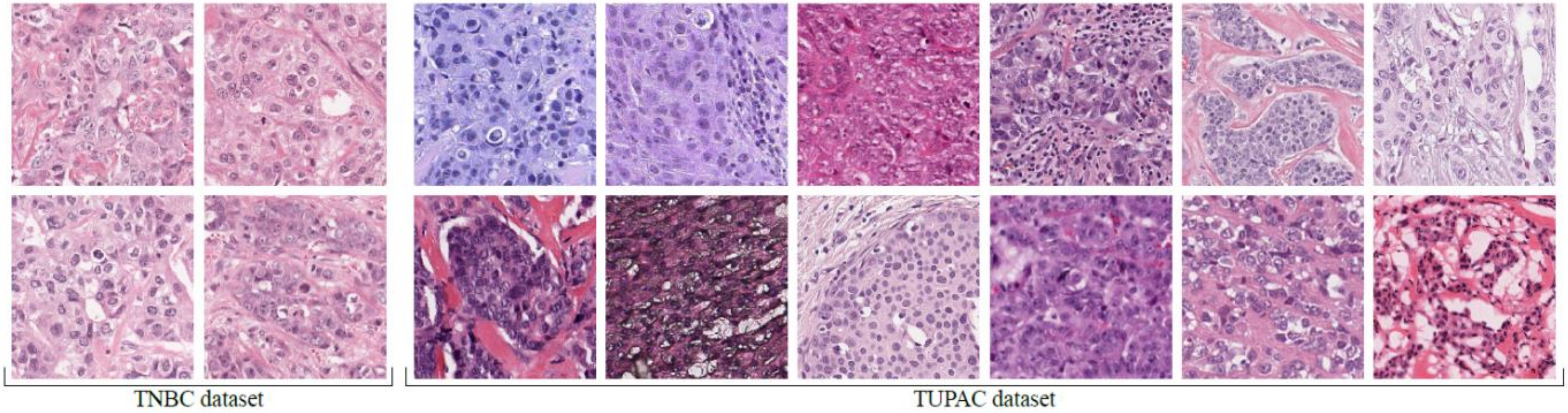

Mitotic detections


Mitosis density

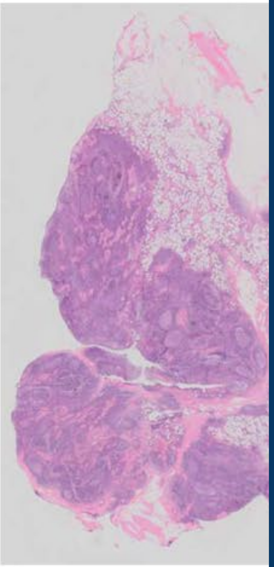

56 mitoses
Direct visibility of hotspots


38 mitoses

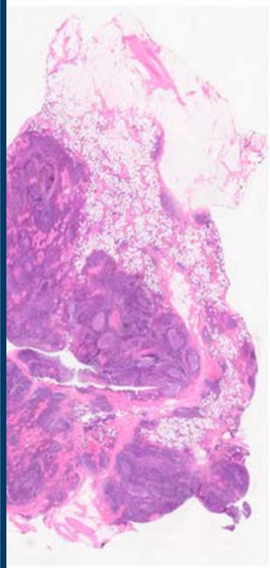
Challenge 2: staining differences



‘Traditional’ stain normalization

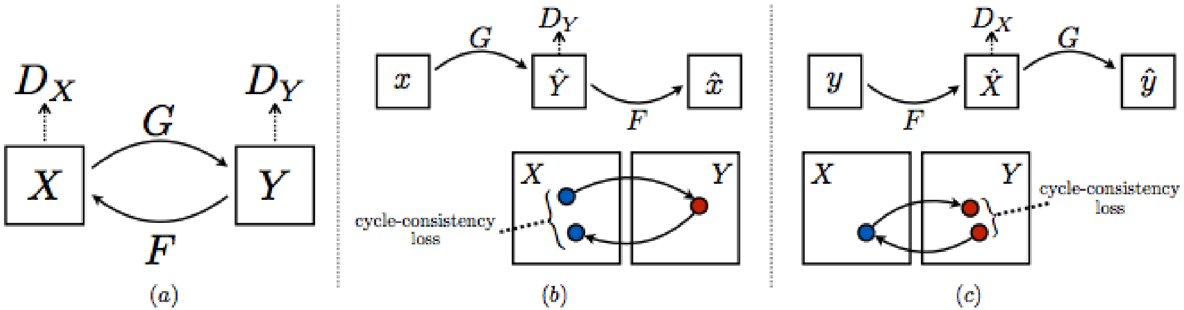


- Only modifies color information
- Very time-consuming algorithm
- Dependent on presence of nuclei
- Parameter tweaking for new datasets



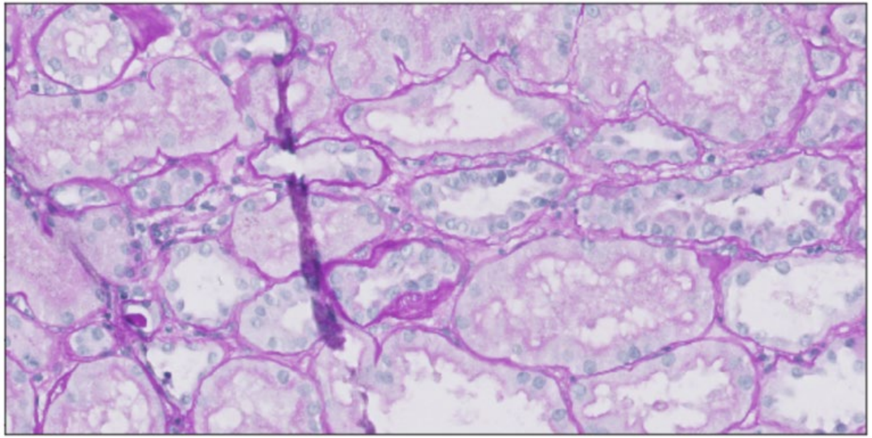
Bejnordi et al. IEEE Trans Med Imag (2016)

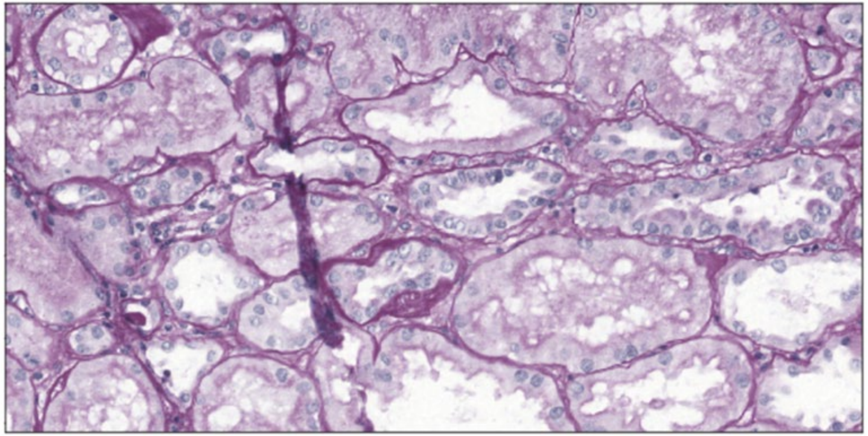
Stain normalization using cycleGANs



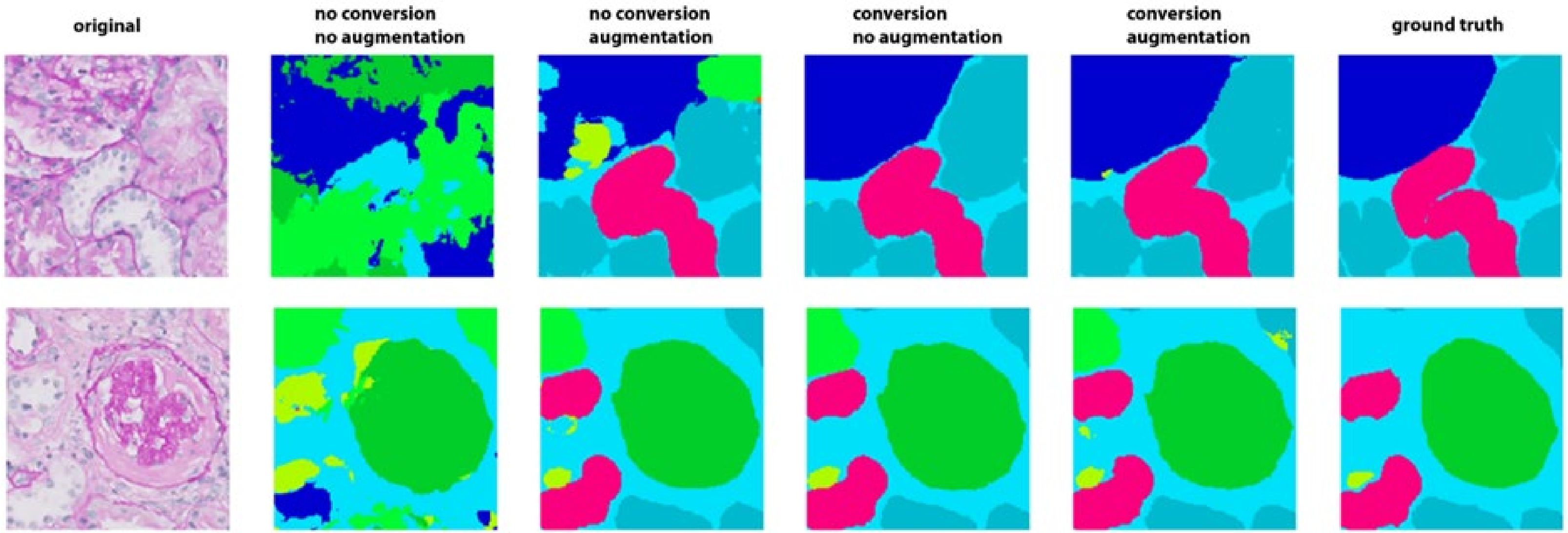
de Bel et al. MIDL (2019)

Stain normalization using cycleGANs





Stain normalization using cycleGANs



Experiment		Dice coefficient AMC			
Augmentations	Stain transformed	Mean	Std	Min	Max
x	x	0.36	0.21	0.09	0.65
x	✓	**0.85**	0.06	0.69	0.91
✓	x	0.78	0.08	0.65	0.87
✓	✓	**0.85**	0.05	0.72	0.91

Practical applications of computation pathology

Detection of metastases in lymph nodes

Automatic mitotic counts

Tumor/stroma ratio quantification

Identific
assoc

Tumor/stroma ratio quantification

Annals of Oncology

original articles

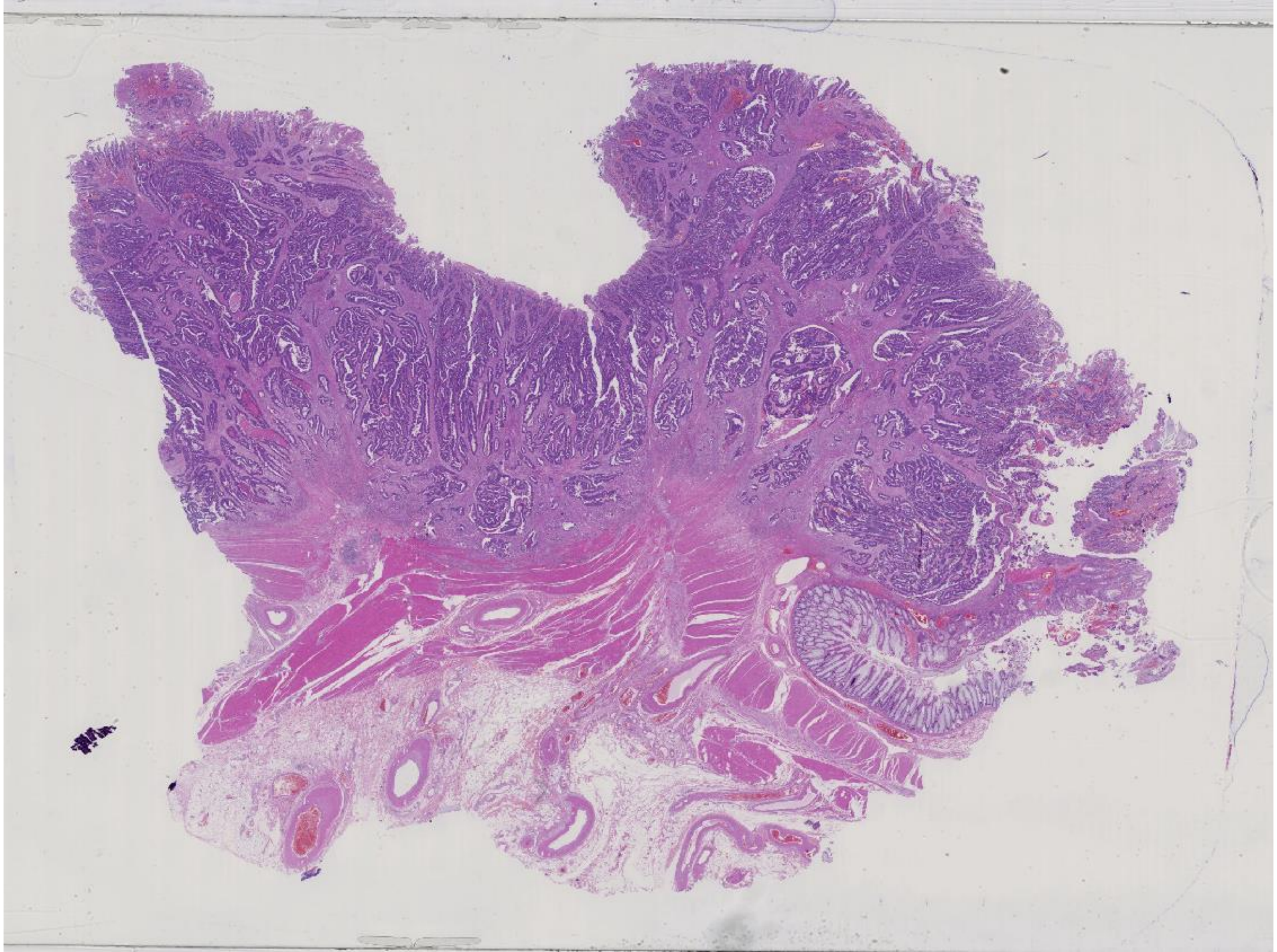
Annals of Oncology 24: 179–185, 2013
doi:10.1093/annonc/mds246
Published online 2 August 2012

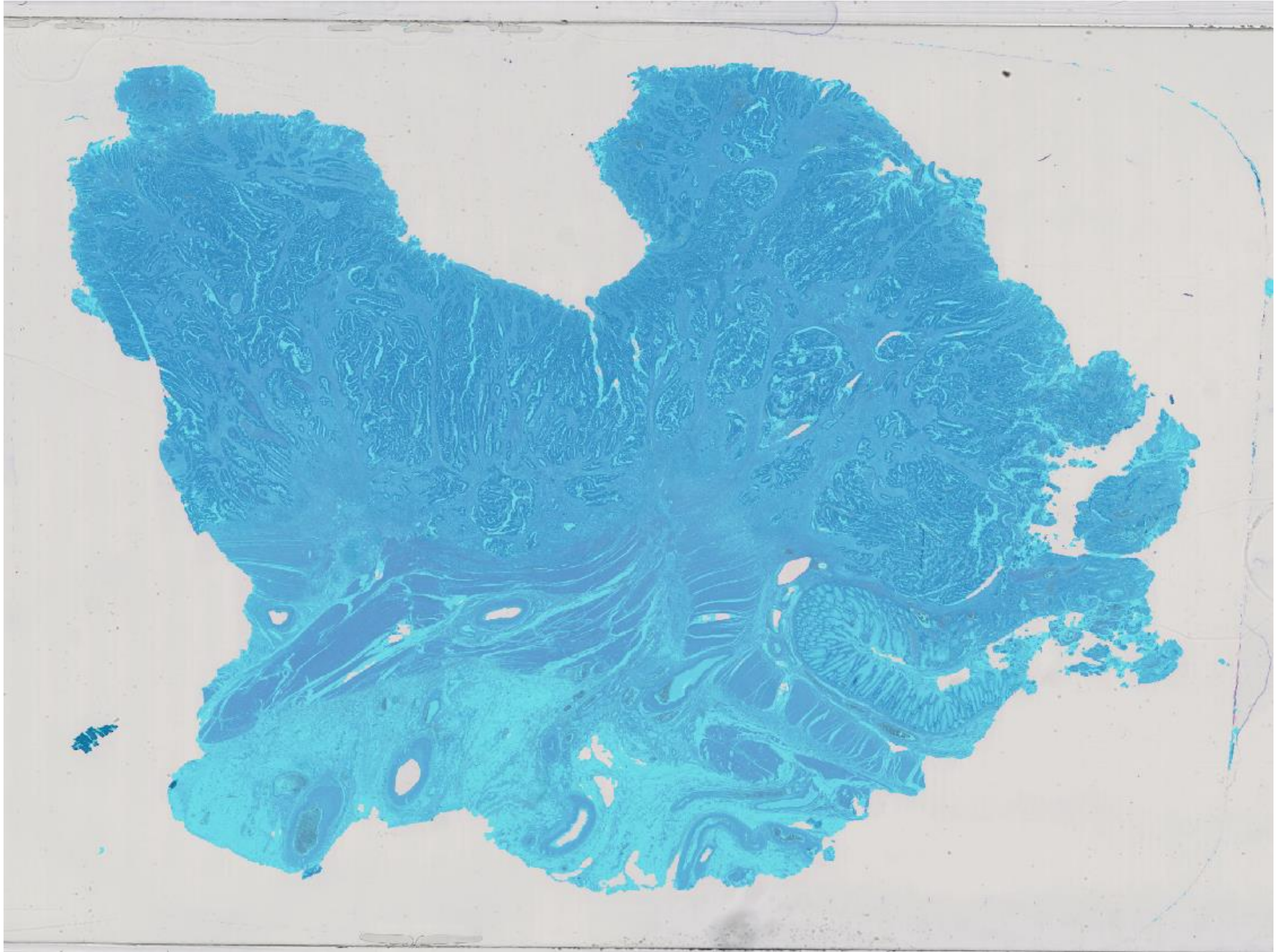
The proportion of tumor-stroma as a strong prognosticator for stage II and III colon cancer patients: validation in the VICTOR trial

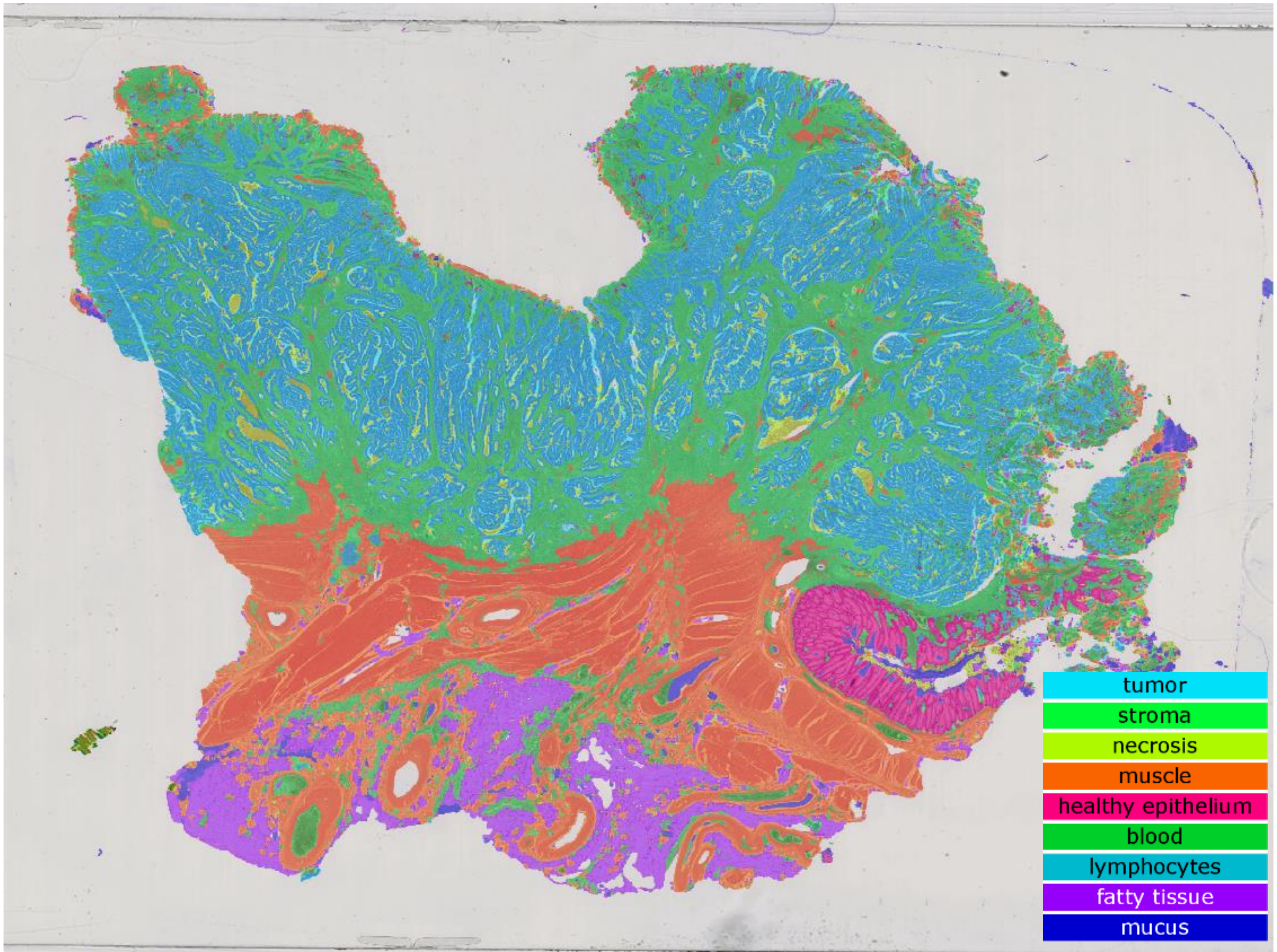
A. Huijbers[1], R. A. E. M. Tollenaar[1], G. W. v Pelt[1], E. C. M. Zeestraten[1], S. Dutton[3], C. C. McConkey[6], E. Domingo[7], V. T. H. B. M. Smit[2], R. Midgley[4], B. F. Warren[8], E. C. Johnstone[4], D. J. Kerr[5] & W. E. Mesker[1]*

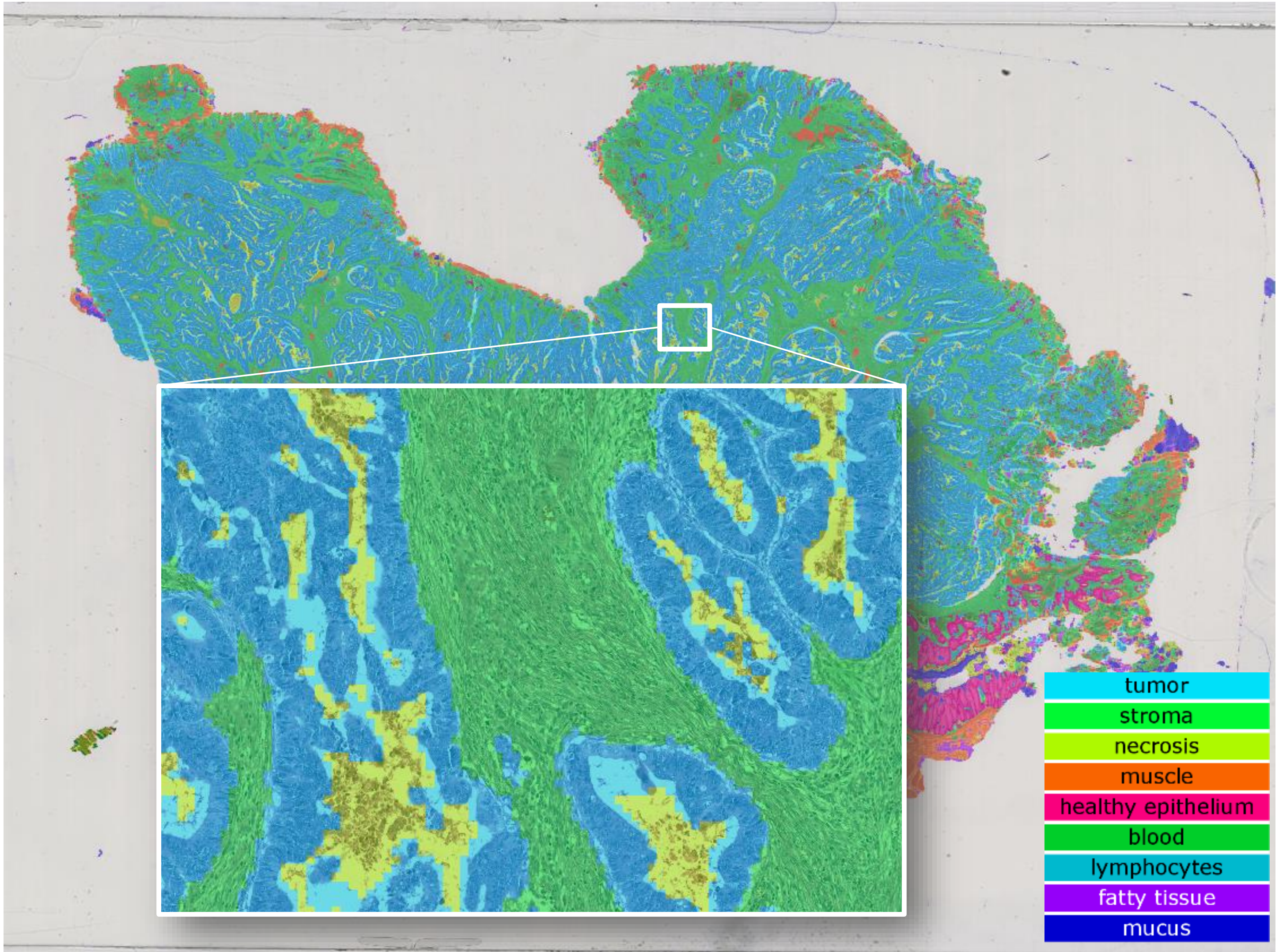
Departments of [1]Surgery; [2]Pathology, Leiden University Medical Center (LUMC), Leiden, The Netherlands; [3]Centre for Statistics in Medicine, University of Oxford, Oxford; Departments of [4]Oncology; [5]Clinical Pharmacology, University of Oxford, Oxford; [6]Clinical Trials Unit, University of Warwick, Coventry; [7]Molecular and Population Genetics, Wellcome Trust Center for Human Genetics, Oxford; [8]Department of Pathology, John Radcliffe Hospital, Headington, Oxford, UK

Received 28 February 2012; revised 15 June 2012; accepted 18 June 2012

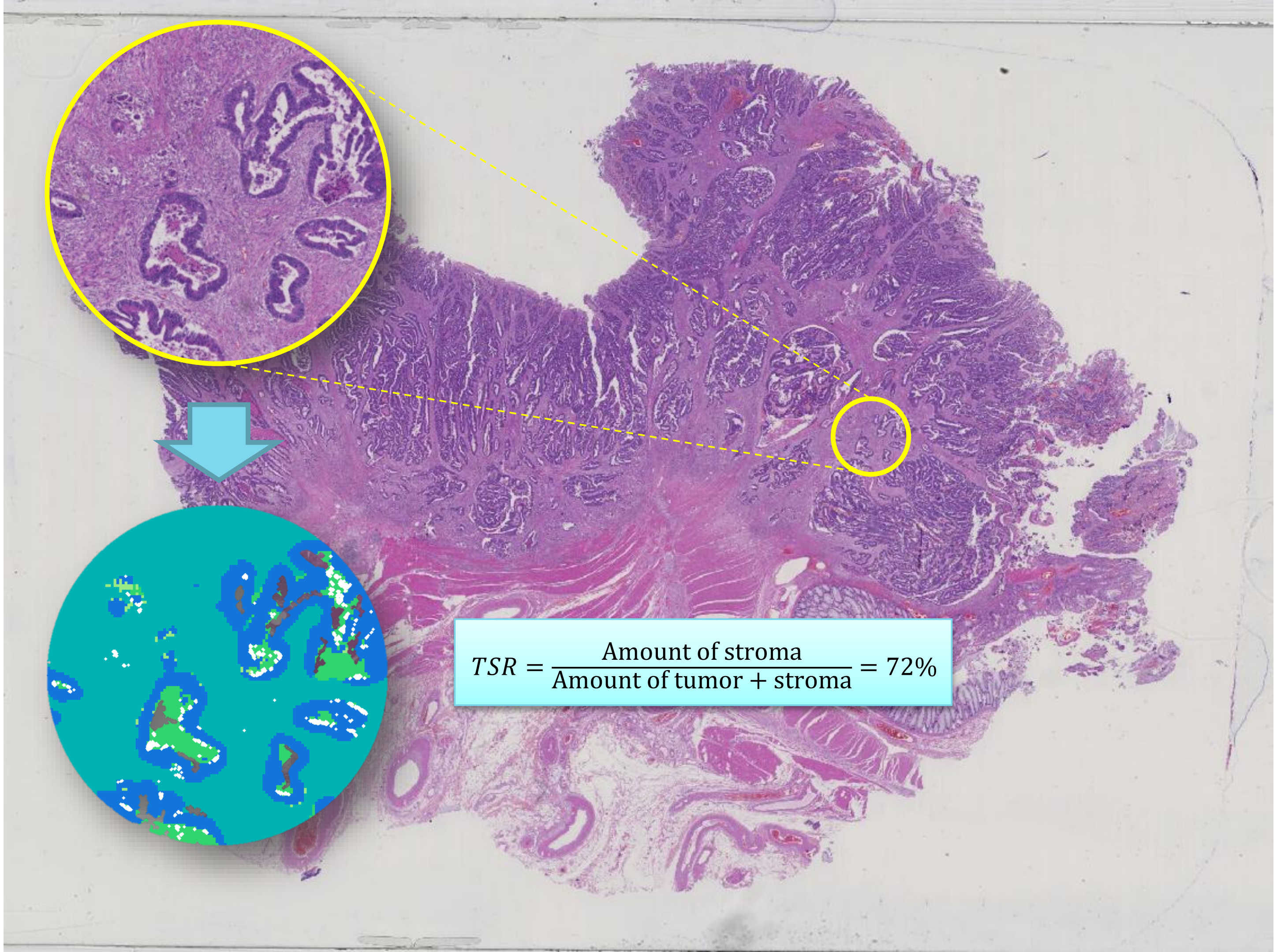





tumor
stroma
necrosis
muscle
healthy epithelium
blood
lymphocytes
fatty tissue
mucus



tumor
stroma
necrosis
muscle
healthy epithelium
blood
lymphocytes
fatty tissue
mucus



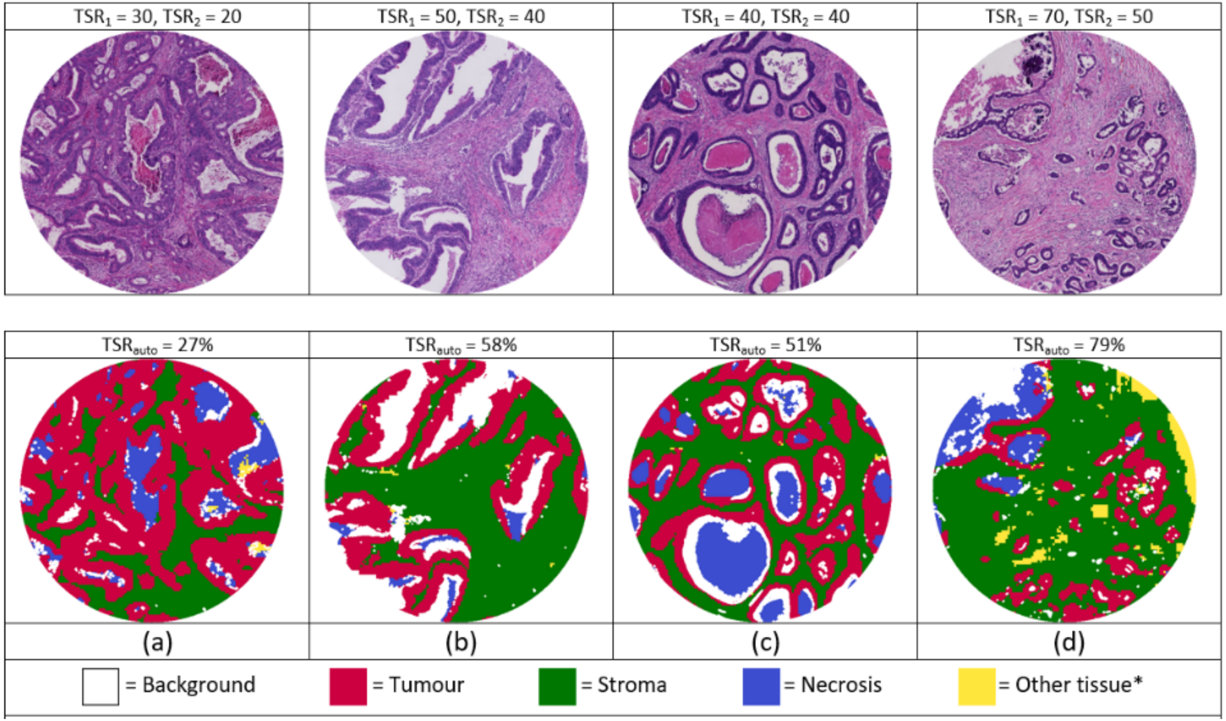
$$TSR = \frac{\text{Amount of stroma}}{\text{Amount of tumor + stroma}} = 72\%$$

Tumor/stroma ratio quantification

125 patients with rectal carcinoma

- Stage I-III
- At least five year follow-up
- No neo-adjuvant therapy

Tumor/stroma ratio quantification



Tumor/stroma ratio quantification

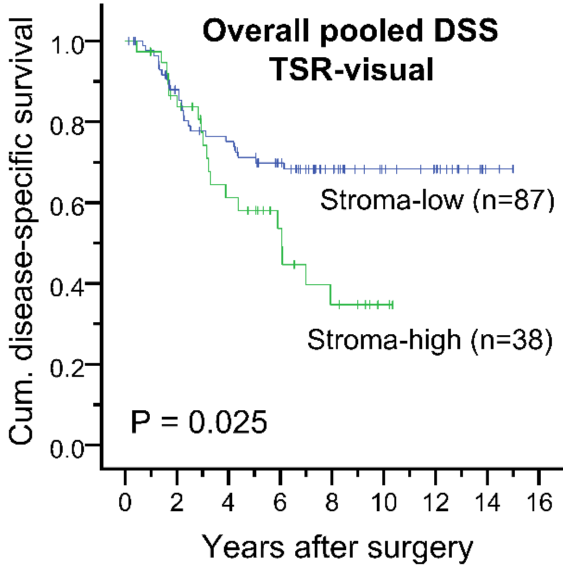
Crosstab: Observer 1 versus Observer 2

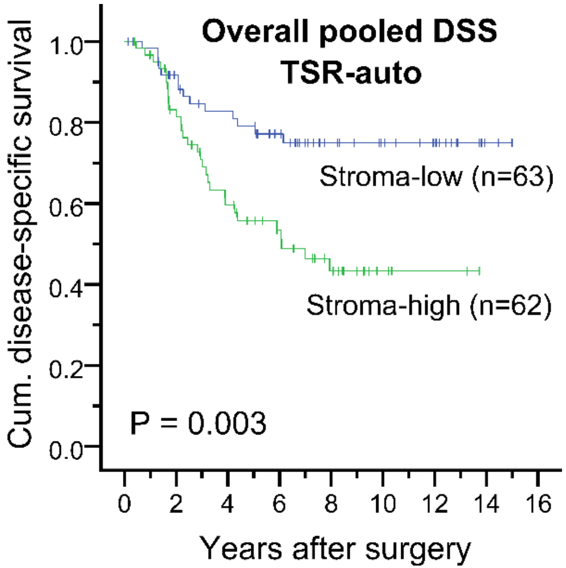
κ = 0.548		Observer 2		
		Stroma-low	Stroma-high	Total
Observer 1	Stroma-low	75	8	83
	Stroma-high	16	26	42
	Total	91	34	125

Crosstab: TSR-Visual (consensus) versus TSR-auto

κ = 0.518		TSR-auto		
		Stroma-low	Stroma-high	Total
TSR-visual (consensus)	Stroma-low	60	27	87
	Stroma-high	3	35	38
	Total	63	62	125

Tumor/stroma ratio quantification

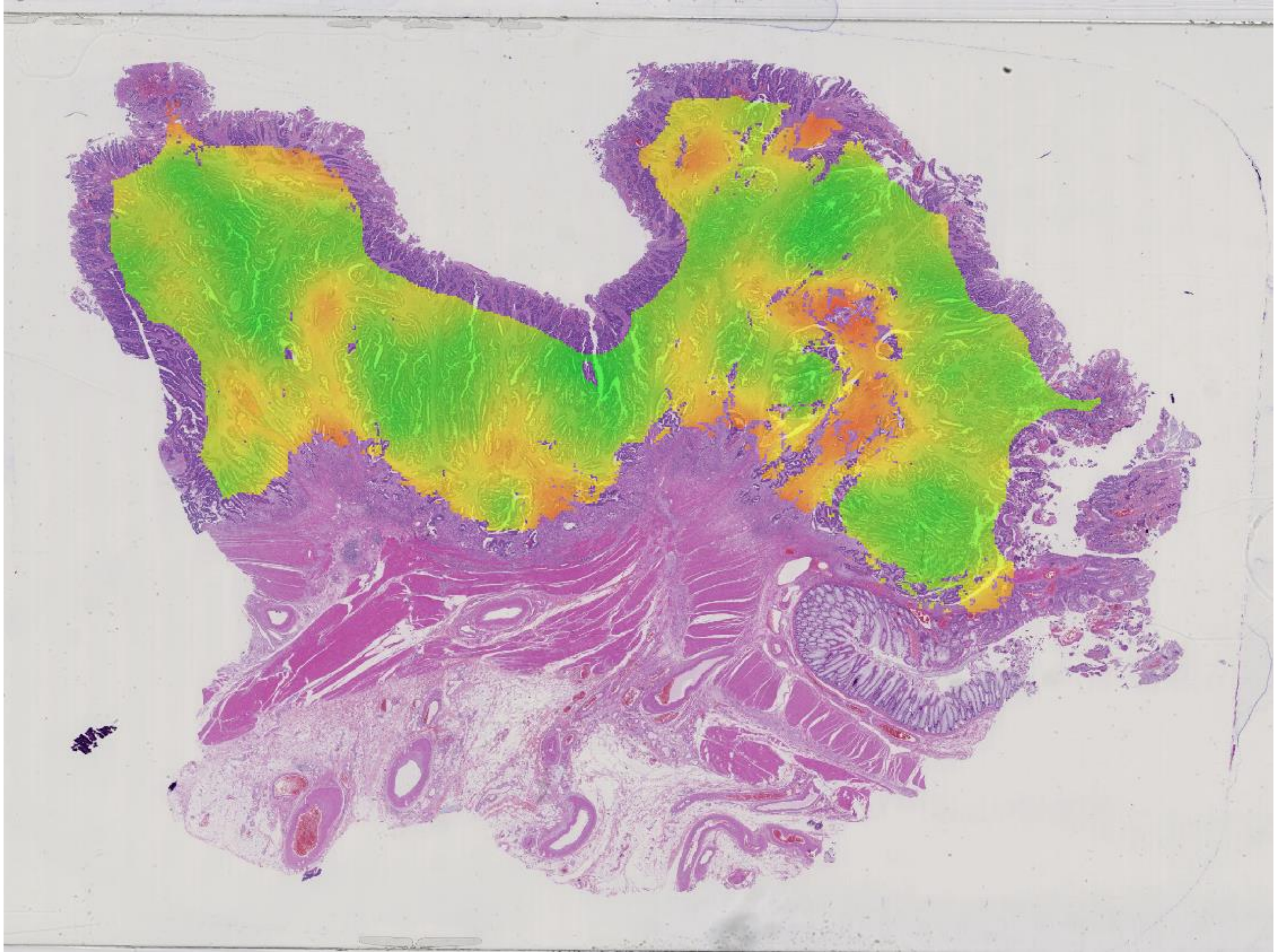




Tumor/stroma ratio quantification

Table 5. Uni- and multivariate Cox regression analysis for disease-specific survival.						
	Univariate analysis		Multivariate analysis			
			Visual		Auto	
	HR (95% CI)	P-val.	HR (95% CI)	P-val.	HR (95% CI)	P-val.
Age	1.01 (0.98-1.04)	0.376				
Gender	0.85 (0.45-1.60)	0.604				
T-stage	2.42 (1.47-3.99)	**0.001**	1.97 (1.16-3.34)	**0.012**	2.05 (1.24-3.38)	**0.005**
N-stage	2.16 (1.49-3.14)	**0.0001**	2.06 (1.13-3.75)	**0.018**	2.12 (1.17-3.84)	**0.014**
Surgical procedure	1.48 (0.94-2.31)	0.090				
Tumour grade	2.96 (1.42-6.17)	**0.004**	2.40 (1.05-5.48)	**0.038**	2.23 (0.99-5.00)	0.052
Adj. chemoth.	1.17 (0.28-4.82)	0.831				
Adj. radioth.	2.56 (1.41-4.63)	**0.002**	0.72 (0.27-1.88)	0.496	0.68 (0.27-1.72)	0.417
TSR-visual	1.96 (1.08-3.58)	**0.027**	2.07 (1.09-3.93)	**0.026**		
TSR-auto	2.57 (1.36-4.86)	**0.004**			2.75 (1.44-5.27)	**0.002**

Geessink et al. *Cellular Oncology (2019).*



Practical applications of computation pathology

n

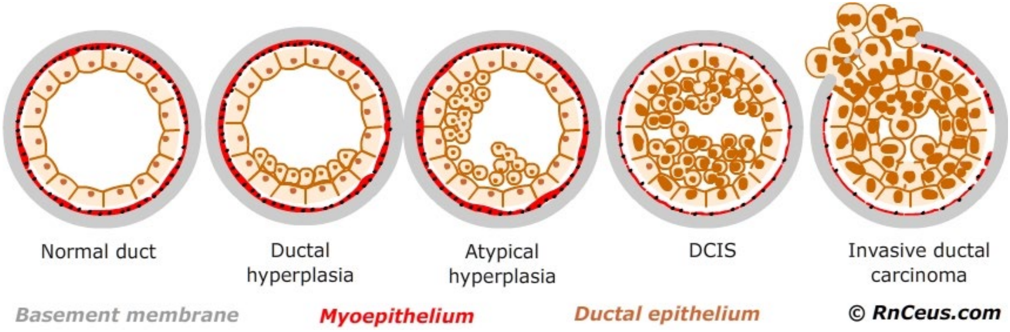
Automatic mitotic counts

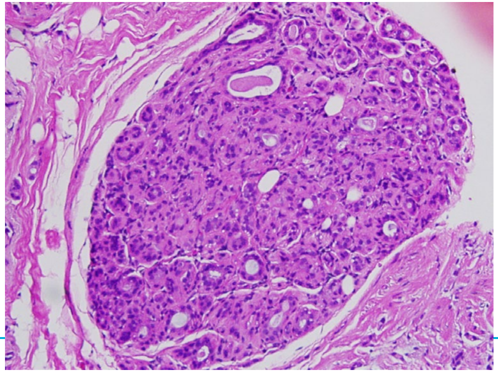
Tumor/stroma ratio quantification

Identification of tumor associated stroma

Pred
ex

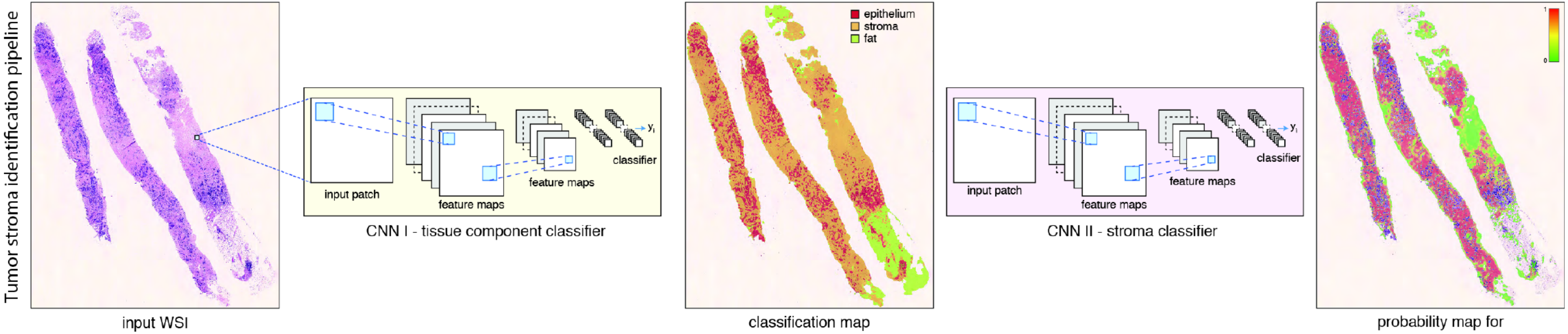
Prognosis of in-situ lesions

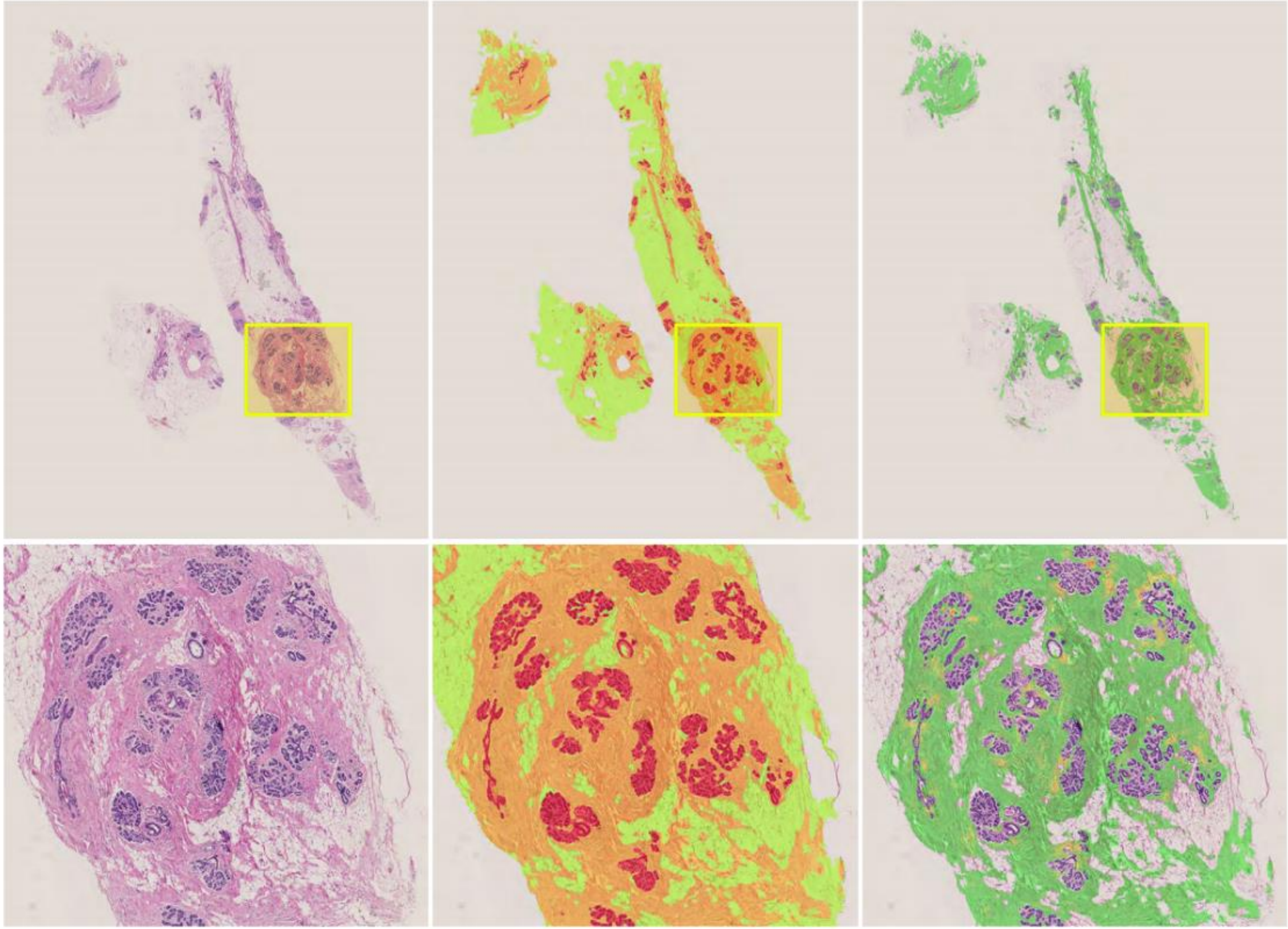


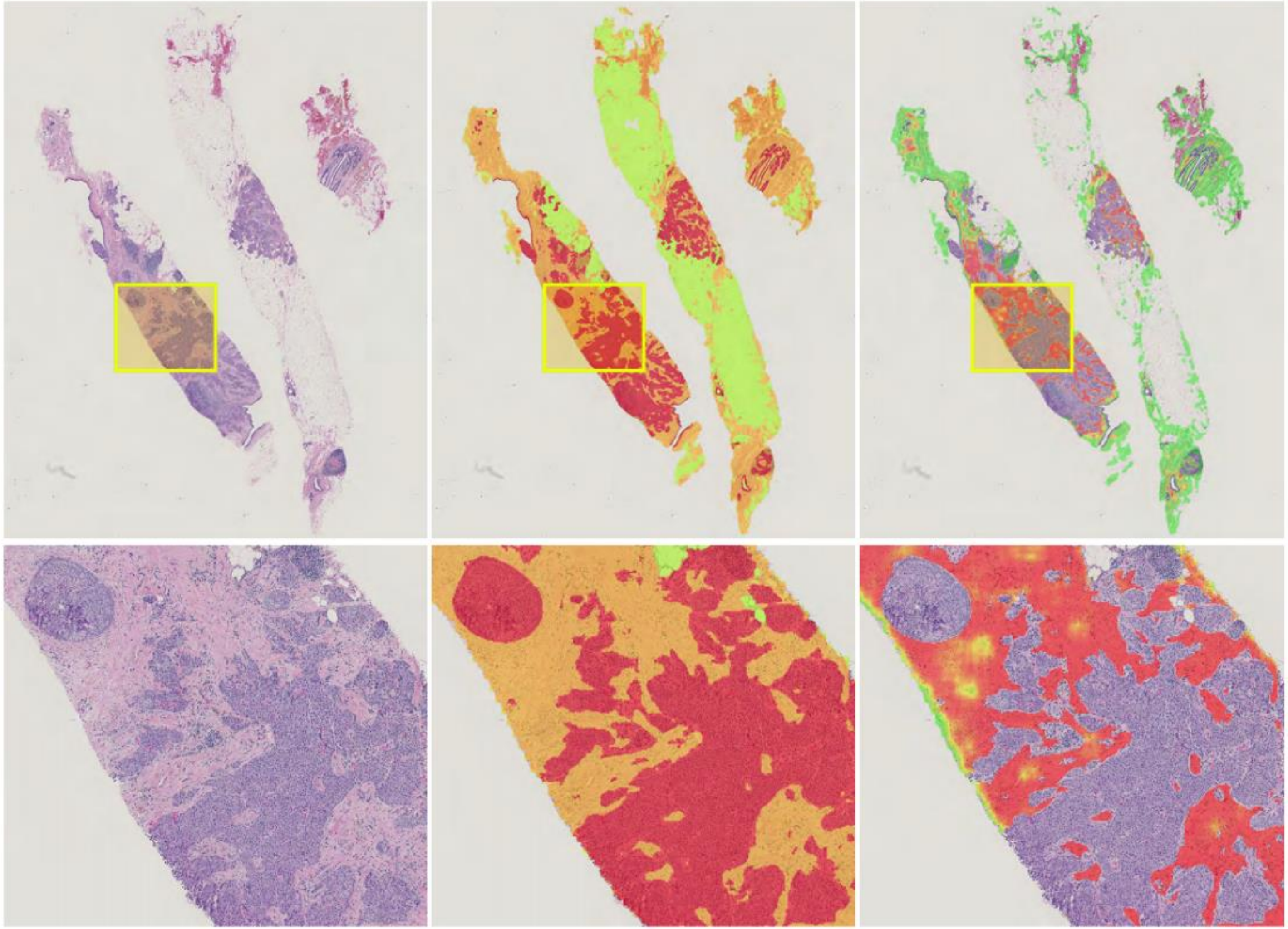


Diagnosis	Entire dataset			Training dataset			Testing dataset		
	# Patient	# WSI	%	# Patient	# WSI	%	# Patient	# WSI	%
Benign	321	675	36.4	209	437	37.9	112	238	33.9
Proliferative	312	937	35.4	209	608	37.9	103	329	32.3
Proliferative with atypia	57	212	6.5	42	171	7.6	15	41	4.5
Ductal carcinoma in-situ	58	222	6.6	–	–	–	58	222	17.6
Lobular carcinoma in-situ	10	29	1.1	7	21	1.2	3	8	0.9
Invasive breast cancer	124	312	14.0	85	222	15.4	39	90	11.8
Total	882	2387	100	552	1459	100	330	928	100

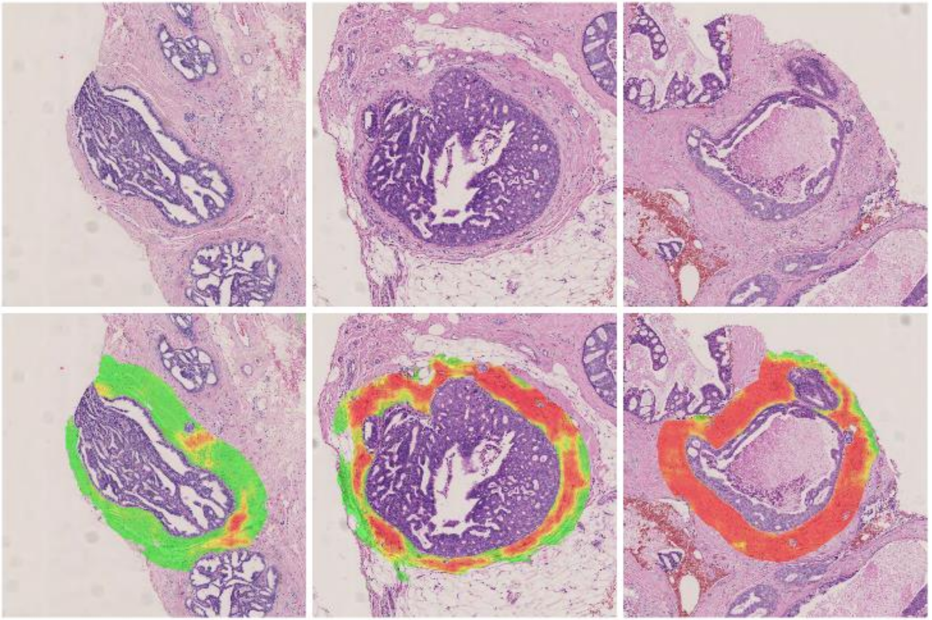
Tumor-associated stroma

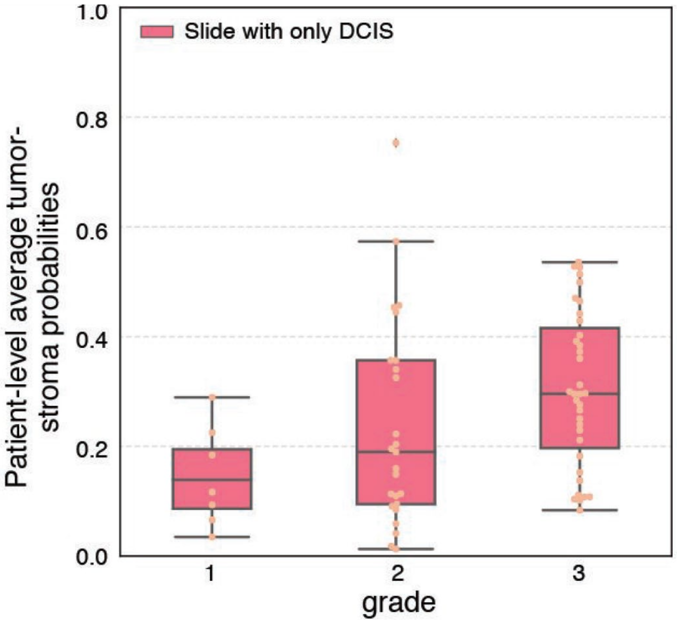






Tumor-associated stroma





Ehteshami et al. *Modern Pathology (2018)*

Practical applications of computation pathology

Tumor/stroma ratio quantification

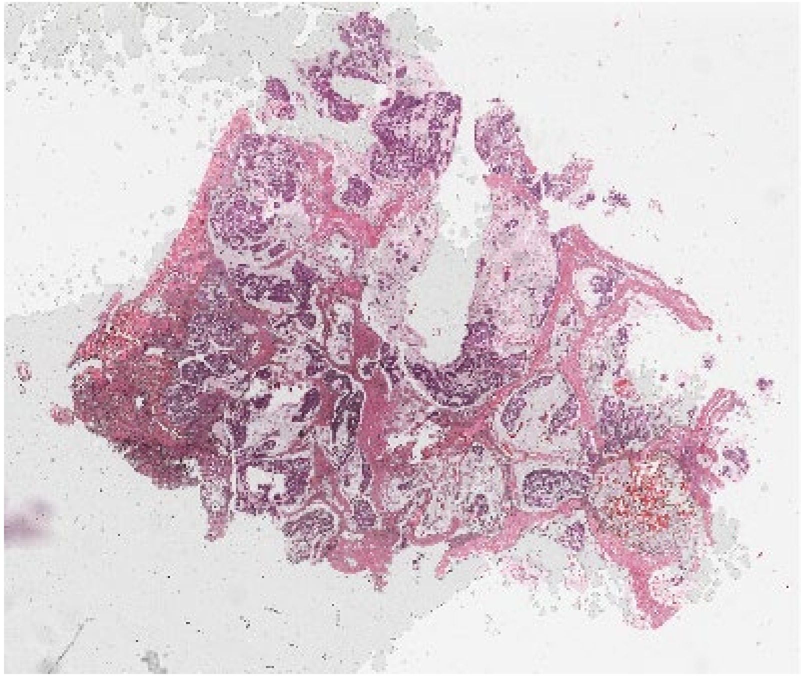
Identification of tumor associated stroma

Prediction of gene expression

Gleas

pros

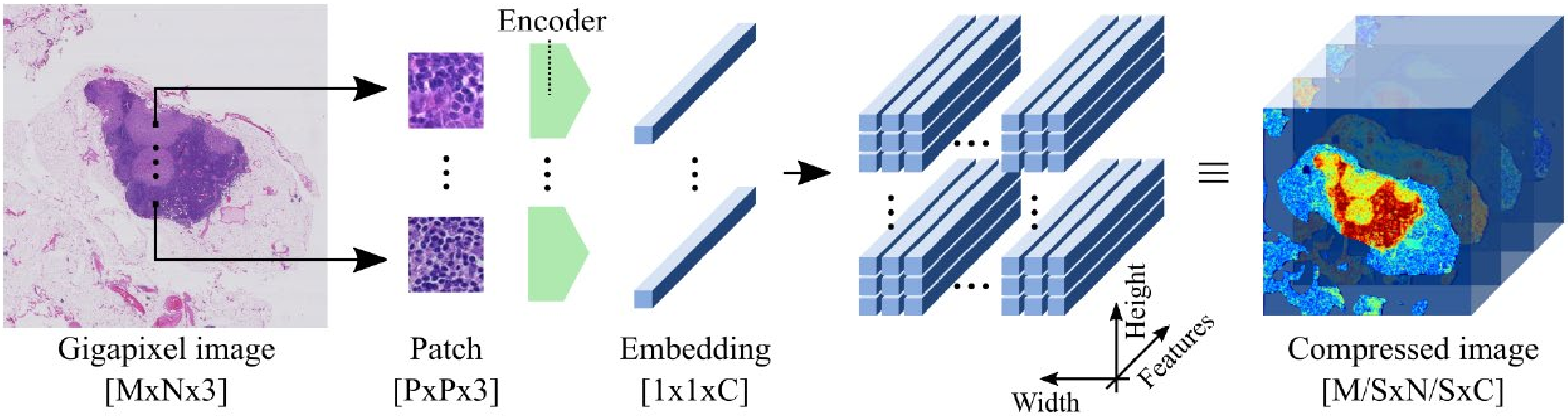
Prediction of gene expression



Mean RNA expression 11 proliferation-associated genes

Score: 0.567

Neural compression

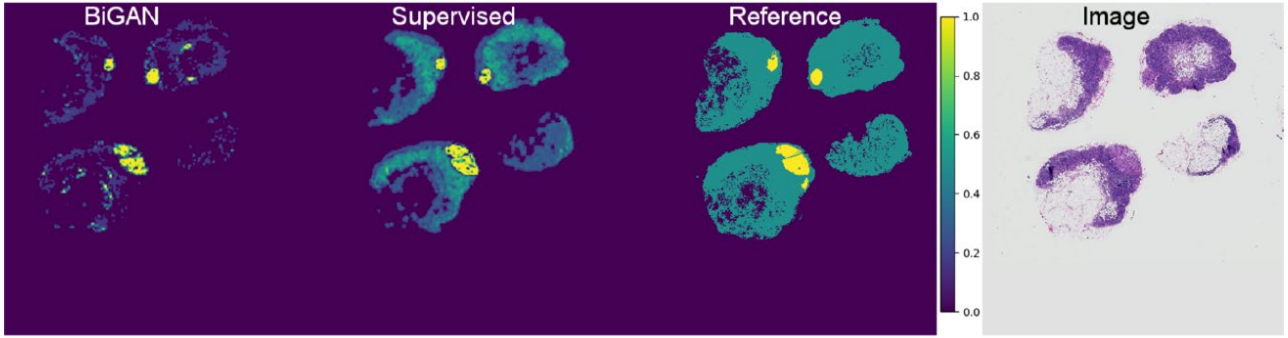


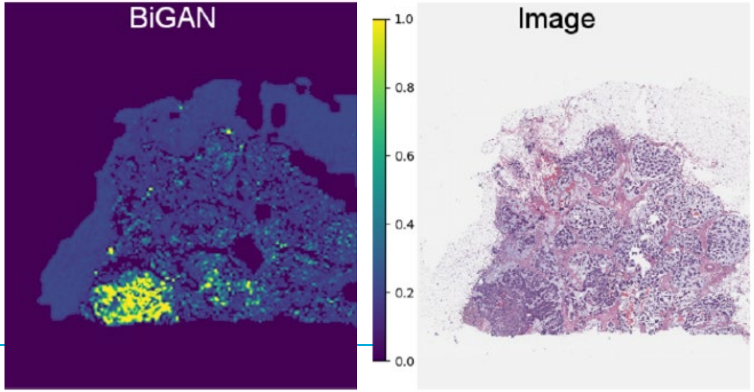
Tellez et al. *In rebuttal*

Prediction of gene expression

Team	Spearman's ρ
Lunit (mitosis counting)	0.617
Radboud (neural compression)	0.557
Radboud (regular CNN)	0.516
ContextVision	0.503

Explainability of ML systems





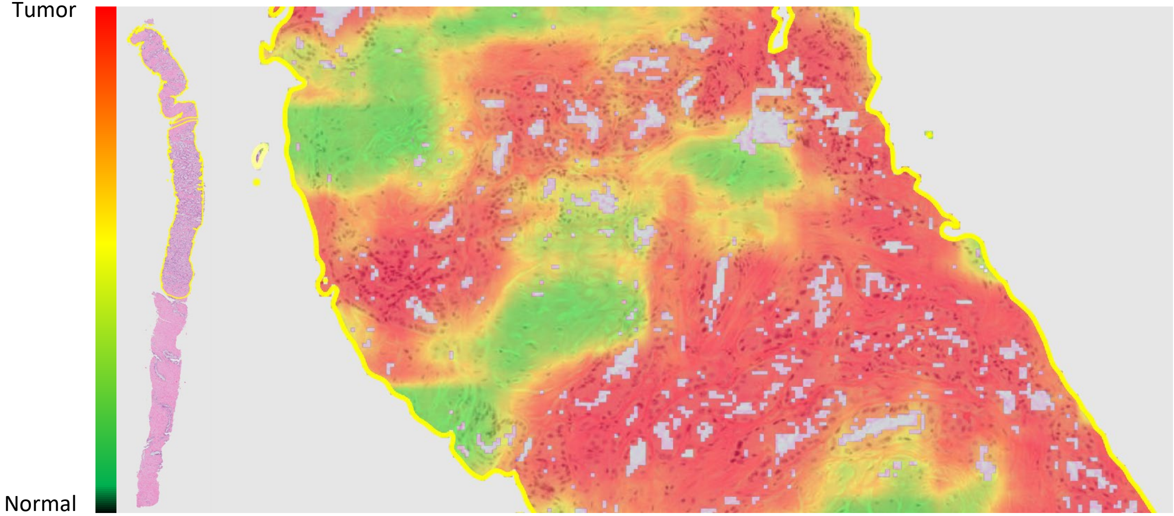
Practical applications of computation pathology

Identification of tumor associated stroma

Prediction of gene expression

Gleason grading of prostate cancer

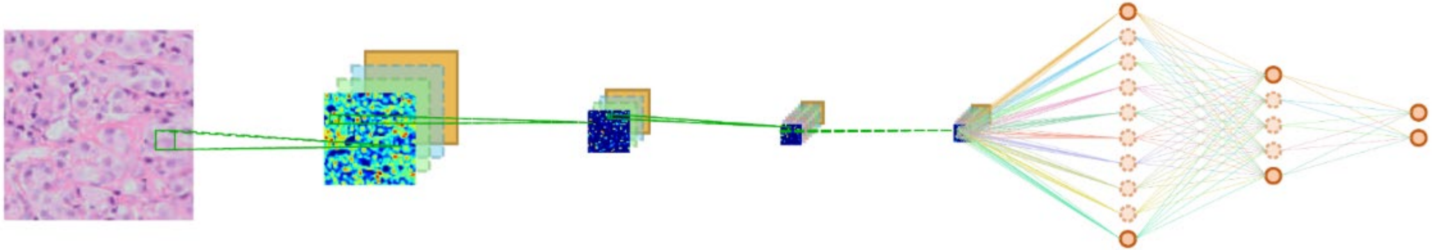
Prostate cancer segmentation



Training set
150 slides
- 83 normal
- 67 cancer

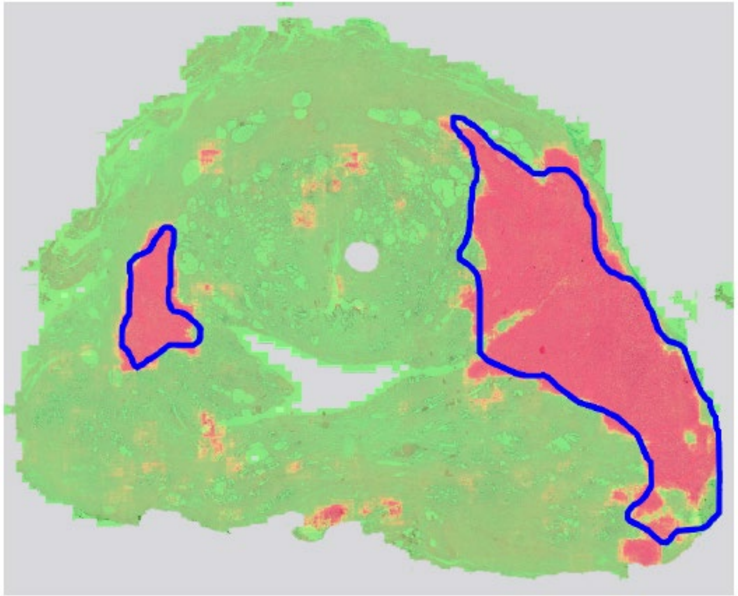
Test set
75 slides
- 30 normal
- 45 cancer

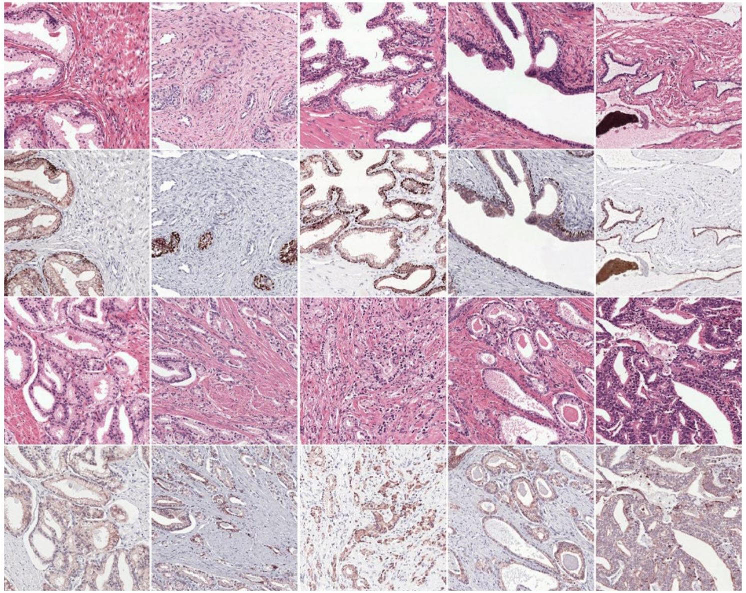
Results
AUC of 0.99...



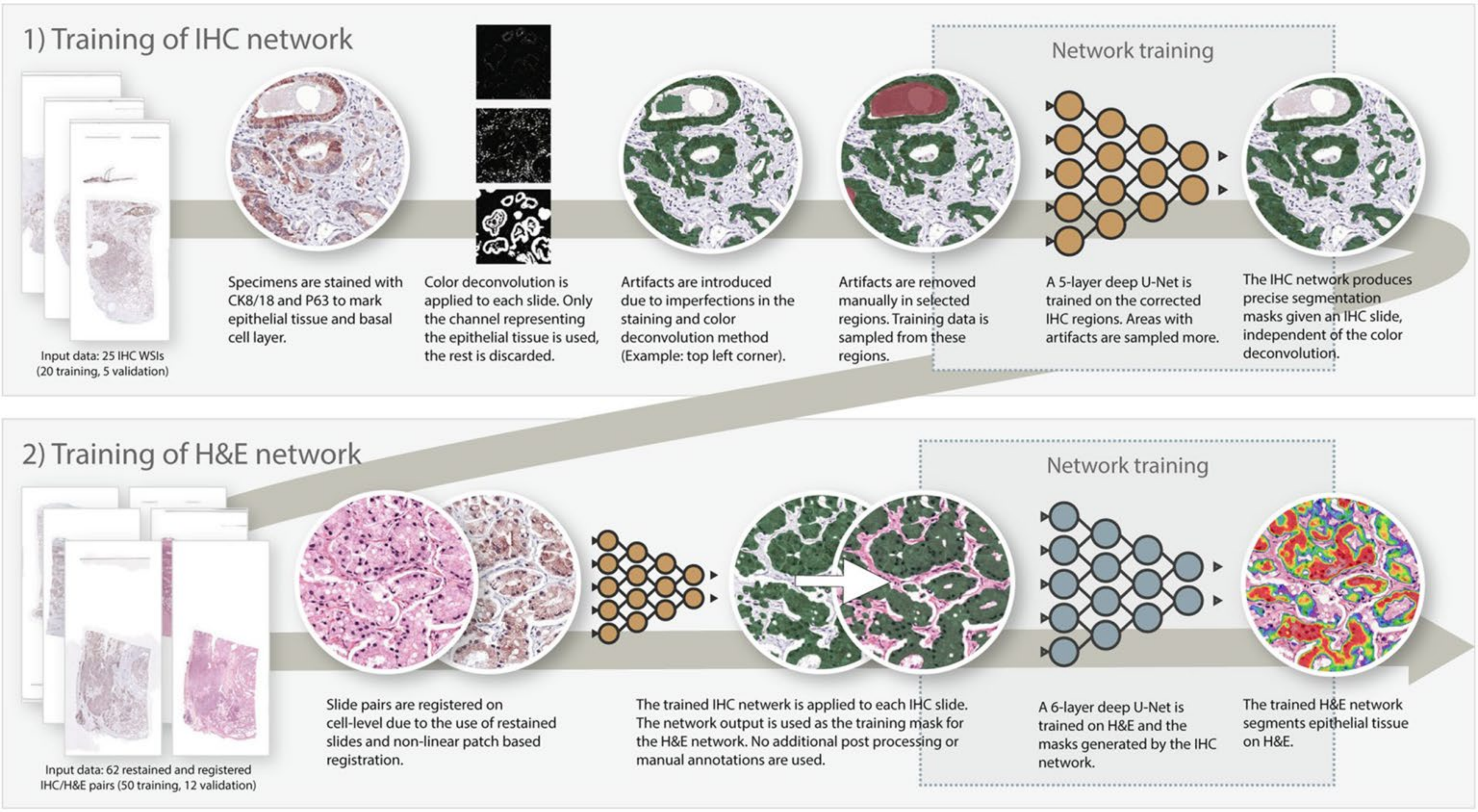
Litjens et al. *Scientific Reports* (2016)

Prostate cancer: epithelium segmentation



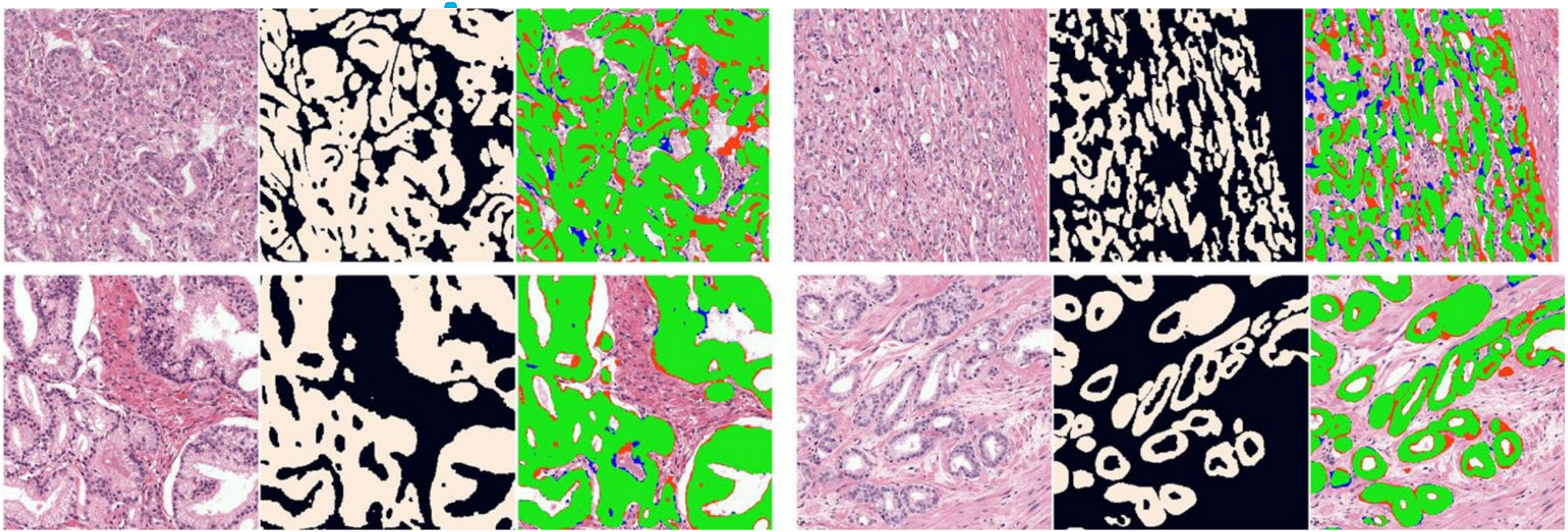


Bulten et al. *Scientific Reports* (2019)



1) Training of IHC network
Network training
Input data: 25 IHC WSIs
(20 training, 5 validation)
Specimens are stained with CK8/18 and P63 to mark epithelial tissue and basal cell layer.
Color deconvolution is applied to each slide. Only the channel representing the epithelial tissue is used, the rest is discarded.
Artifacts are introduced due to imperfections in the staining and color deconvolution method (Example: top left corner).
Artifacts are removed manually in selected regions. Training data is sampled from these regions.
A 5-layer deep U-Net is trained on the corrected IHC regions. Areas with artifacts are sampled more.
The IHC network produces precise segmentation masks given an IHC slide, independent of the color deconvolution.
2) Training of H&E network
Network training
Input data: 62 restained and registered
IHC/H&E pairs (50 training, 12 validation)
Slide pairs are registered on cell-level due to the use of restained slides and non-linear patch based registration.
The trained IHC netwerk is applied to each IHC slide. The network output is used as the training mask for the H&E network. No additional post processing or manual annotations are used.
A 6-layer deep U-Net is trained on H&E and the masks generated by the IHC network.
The trained H&E network segments epithelial tissue on H&E.

Prostate cancer: epithelium

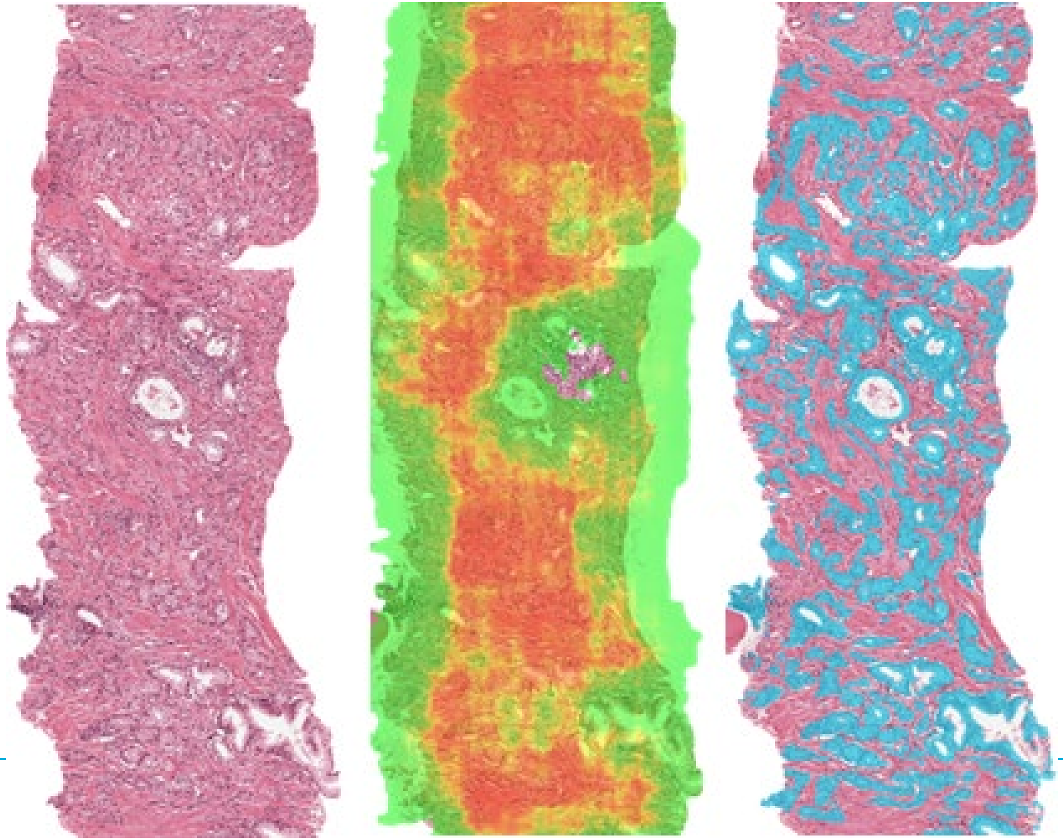


Regions	N	F1 score mean (min, max)	Accuracy	Jaccard
All regions	160	0.893 ± 0.05 (0.661, 0.959)	0.940	0.811

Network	Evaluation	Accuracy	F1	Jaccard
Gertych *et al.*[8]	Cross-validation	—	—	0.595 ± 0.15
Li *et al.*[12]	Cross-validation	—	—	0.737*
Our method	Hold-out validation	0.866 ± 0.07	0.835 ± 0.13	0.735 ± 0.16

Bulten et al. *Scientific Reports* (2019)

Prostate cancer: Gleason Grading

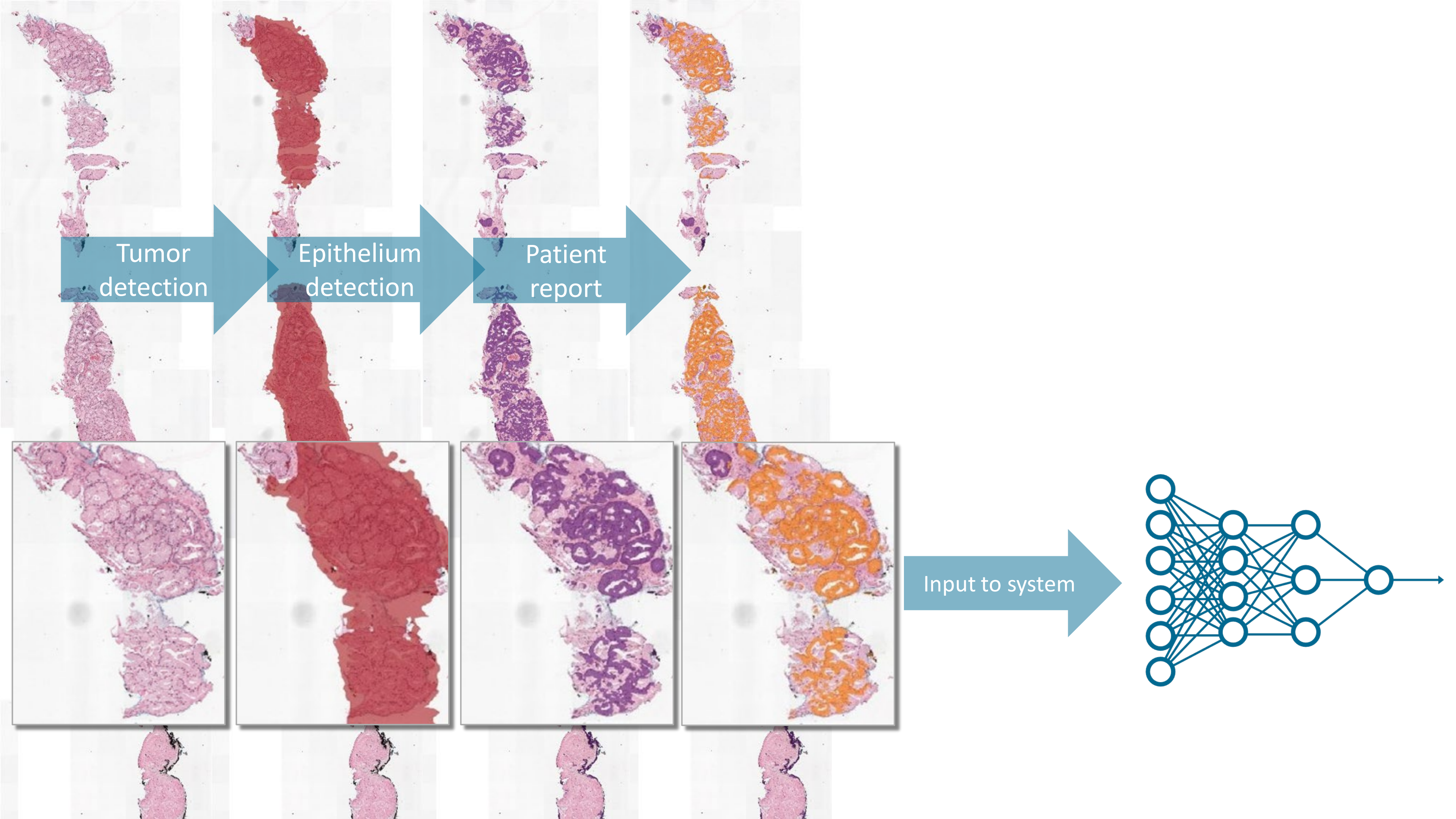


Radboudumc

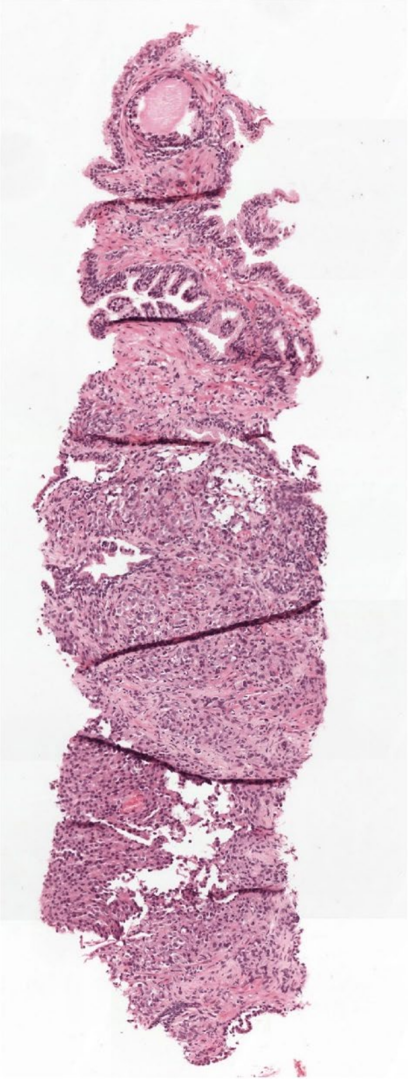
Prostate cancer: Gleason Grading

Collected prostate biopsies from 1271 patients

Grade	Training Set	Validation Set	Test Set
No cancer	777	200	271
3	1508	139	120
4	2102	138	134
5	329	42	100
Totals	4716	519	625



Tumor detection
Epithelium detection
Patient report
Input to system



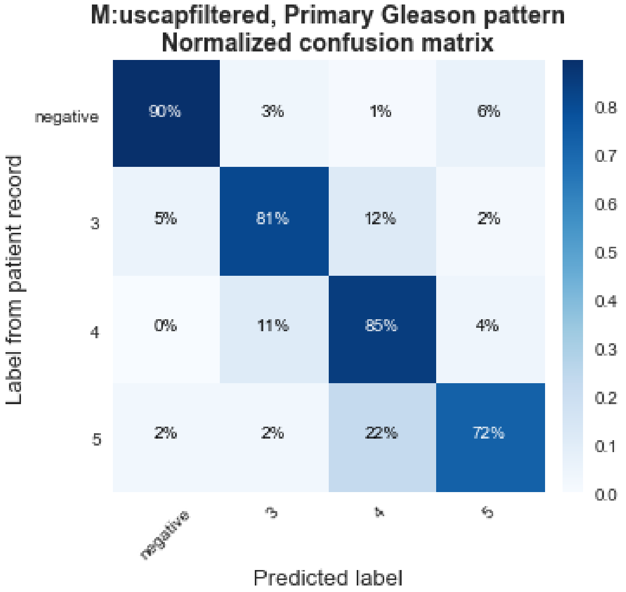
Count

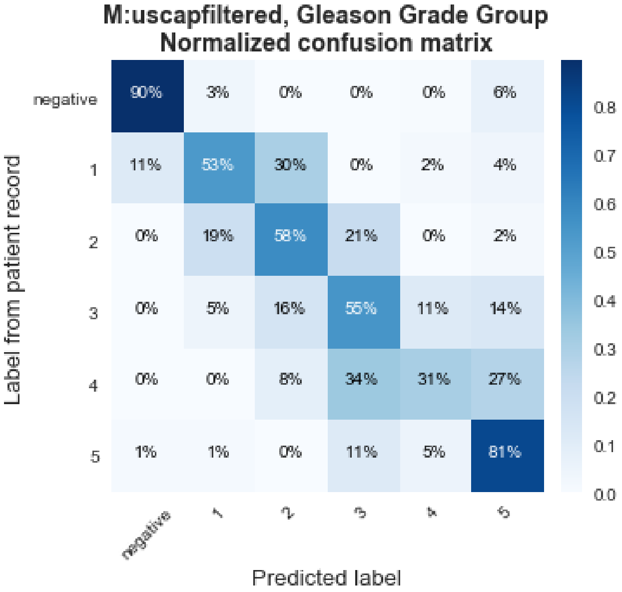
20% benign
15% Gleason 4
65% Gleason 5

Gleason 5+4

Radboudumc

Gleason Grading

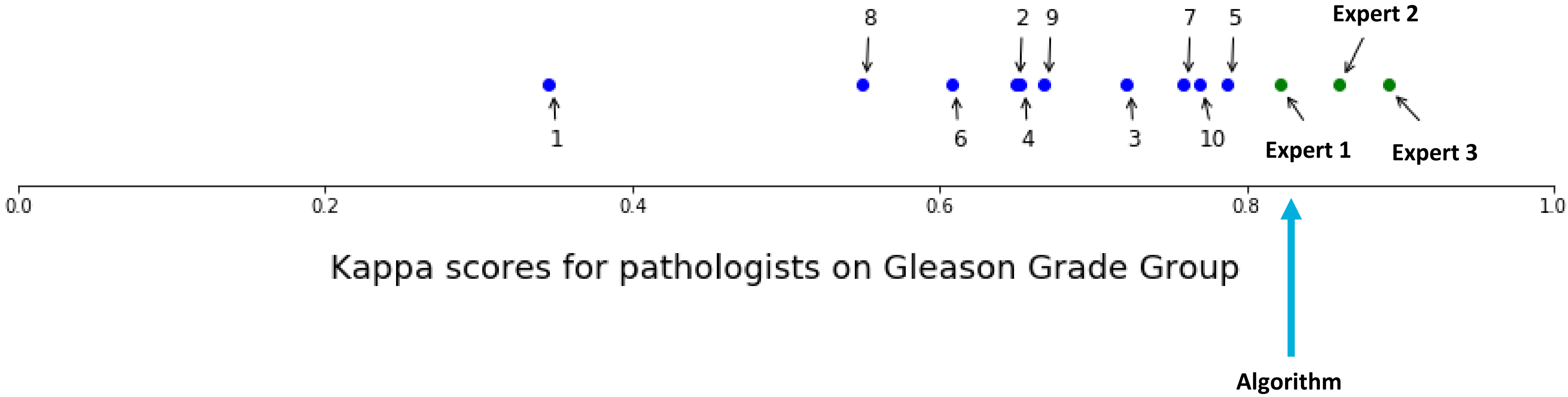




Performance of model on GGG: acc 0.84, k 0.83

Radboudumc

Observer experiment



The people who do all the work…

Scientific staff


Caner Mercan
Postdoctoral researcher


David Tellez
PhD student


Elke Loskamp-Huntink
Study manager


Hans Pinckaers
PhD student


Jasper Linmans
PhD student


John-Melle Bokhorst
PhD student


Mart van Rijthoven
PhD student


Maschenka Balkenhol
Pathology resident and PhD student


Meyke Hermsen
PhD student


Oscar Geessink
PhD student


Péter Bándi
PhD student


Thomas de Bel
PhD student


Wouter Bulten
PhD student


Yiping Jiao
PhD student


Zaneta Swiderska-Chadaj
Postdoctoral researcher

Technical staff


Karel Gerbrands
Research Software Engineer


Maud Wekking
Research technician


Merijn van Erp
Scientific programmer

Visiting researchers


Emiel Stoelinga
Master student


Koen Dercksen
Master student


Leander van Eekelen
Master student


Michel Kok
Master student


Patrick Sonsma
Master student

Faculty


Jeroen van der Laak
Associate professor/Group leader


Geert Litjens
Assistant professor


Francesco Ciompi
Assistant Professor

ST IT PRO
stichting IT projecten

NWO

SEVENTH FRAMEWORK PROGRAMME

Alpe d'HuZes
KWF KANKER BESTRIJDING

ZonMw

RESEARCH & INNOVATION
Horizon 2020

Radboudumc
AMI
Fraunhofer MEVIS

KWF KANKER BESTRIJDING

tEPIS
trait
a ctmm project

Institute for Health Sciences
Radboudumc

Radboudumc | Diagnostic Image Analysis Group

Computational Pathology Group

Highlights Members Projects Software

Computational Pathology Group

The Computational Pathology Group develops, validates and deploys novel medical image analysis methods based on deep learning technology and focusing on computer-aided diagnosis. Application areas include diagnostics and prognostics of breast, prostate and colon cancer. We have rapidly expanded over the last few years, counting over 15 people today. Our group is among the international front runners in the field, witnessed for instance by our highly successful CAMELYON challenges. We have a strong translational focus, facilitated by our close collaboration with clinicians and industry.

Automated tumor detection

computationalpathology.eu

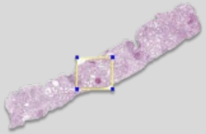
Automating kidney diagnostics

- Glomerular counting
- *ct* score vs % Atrophic tubuli

Quantitative criteria for tubular atrophy: *ct* score	
ct0	No tubular atrophy
ct1	Tubular atrophy involving up to 25% of the area of cortical tubules (mild tubular atrophy)
ct2	Tubular atrophy involving up to 26-50% of the area of cortical tubules (moderate tubular atrophy)
ct3	Tubular atrophy involving in >50% of the area of cortical tubules (severe tubular atrophy)

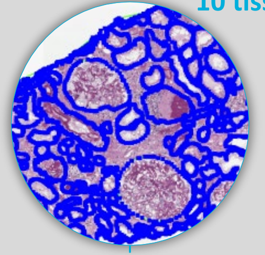
Hermsen et al. *In rebuttal* (2019)

Automating kidney diagnostics

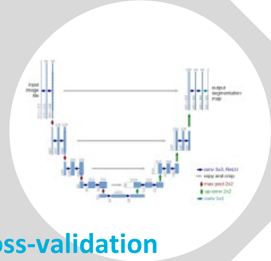


40 fully annotated regions

Checked by pathologist

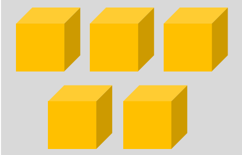


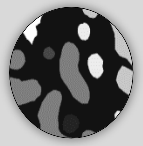
10 tissue classes



5-fold cross-validation

U-net ensemble

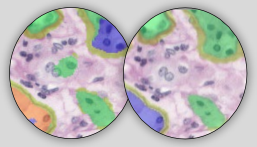




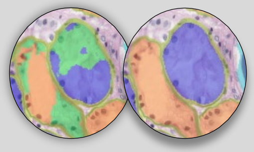
Un-annotated tissue inside region assigned to ‘Interstitium’

Extraction of 4 pixels of all masks for ‘Border’ annotations

Post-processing



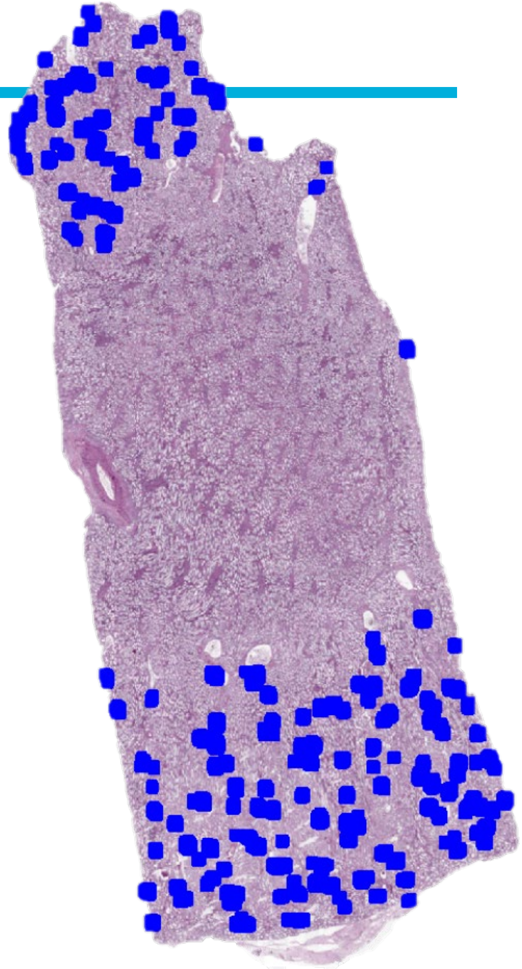
Removal of objects <300 pixels



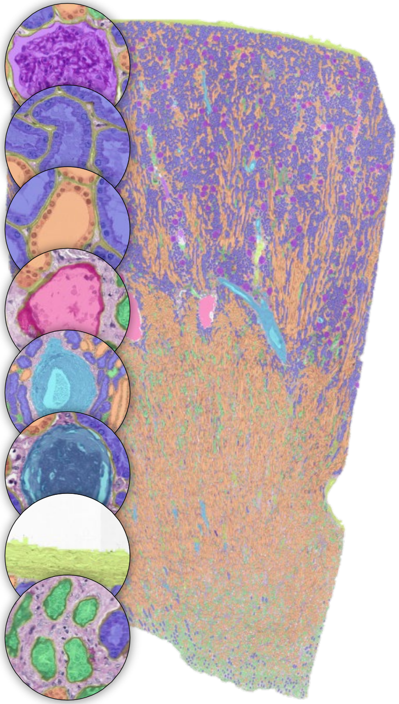
1 class per detected object

Automating kidney diagnostics

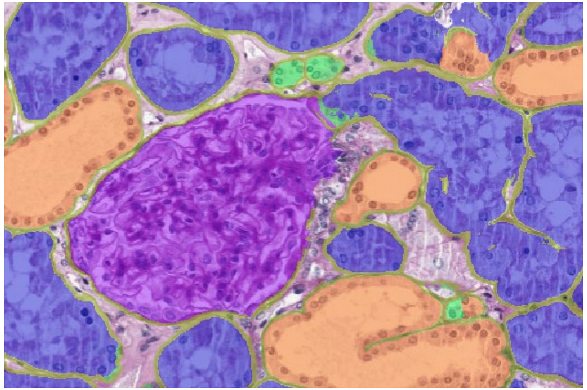
- Applied to 15 WSIs of large tumor nephrectomies
- All glomeruli annotated
    - *1747 Glomeruli and 72 Sclerotic glomeruli*



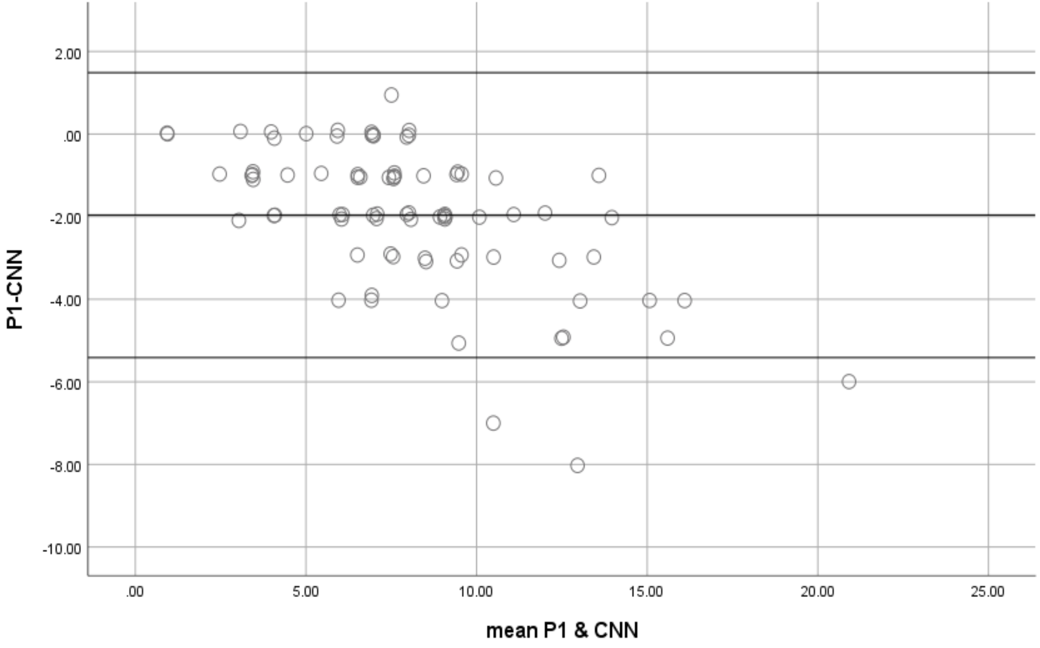
Automating kidney diagnostics



	TP	FP	FN
Glomeruli (n=1747)	93.4% (1632)	8.4 % (149)	6.6 % (115)
Sclerotic glomeruli (n=76)	76.4 % (55)	45.5 % (46)	23.6 % (17)
Total (n=1819)	**92.7 % (1687)**	**10.4 % (192)**	**7.3 % (132)**



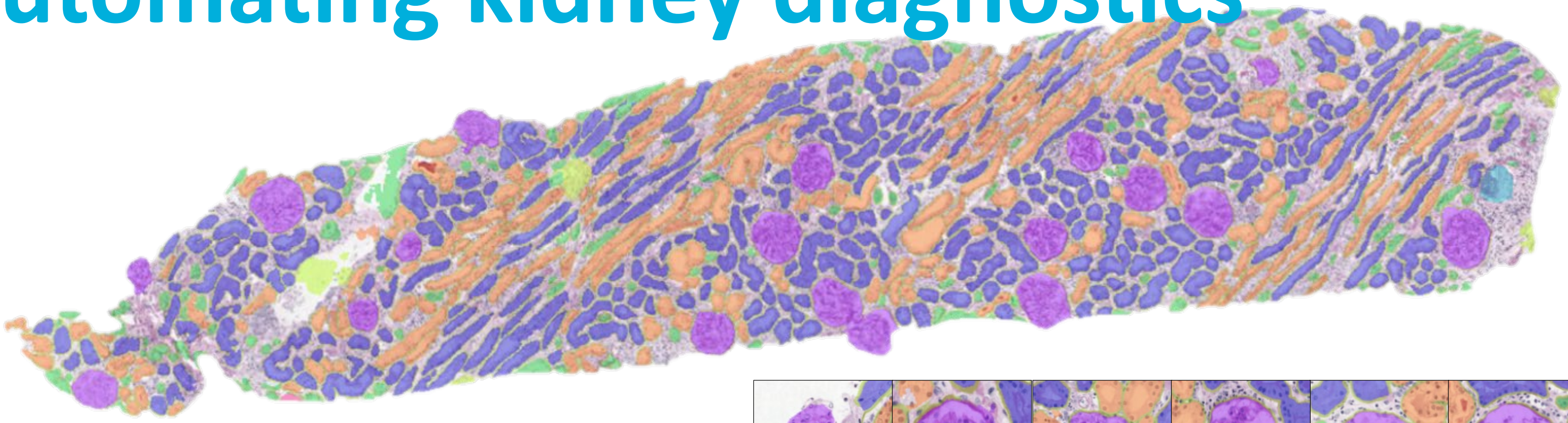
Automating kidney diagnostics



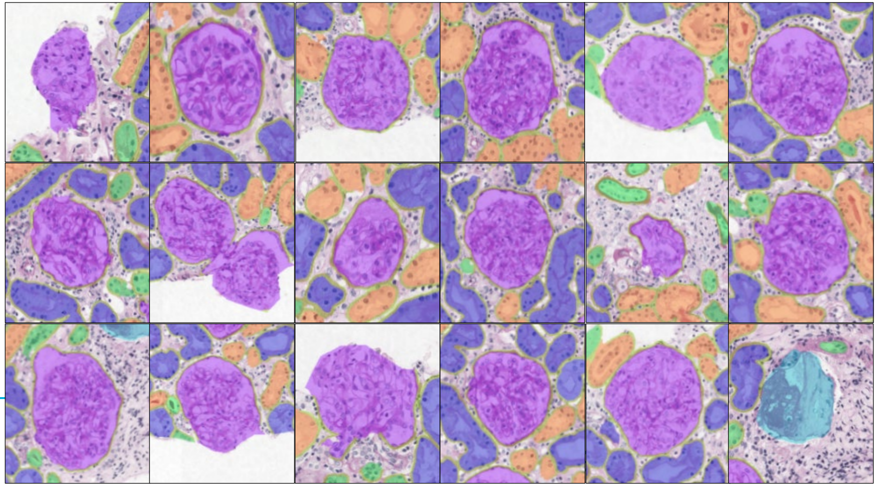
Inter-class correlation coefficient

	P1	P2	P3	CNN
P1		0.94	0.95	**0.78**
P2			0.95	**0.85**
P3				**0.85**
CNN				

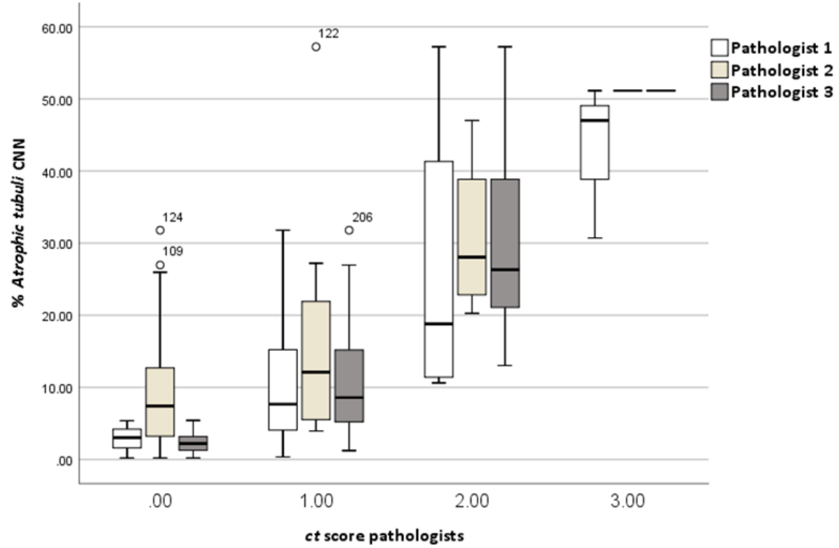
Automating kidney diagnostics



	No.
Pathologist 1	13
Pathologist 2	13
Pathologist 3	14
CNN *Glomeruli*	17
CNN *Sclerotic glomeruli*	1



Automating kidney diagnostics



Bonferroni analysis

	0	1	2	3
0		0.24	<0.001	<0,001
1			<0.001	<0.001
2				<0.01
3				

Weighted kappa

	P1	P2	P3
P1		0.13	0.34
P2			0.20
P3			

Quantitative criteria for tubular atrophy: *ct* score	
ct0	No tubular atrophy
ct1	Tubular atrophy involving up to 25% of the area of cortical tubules (mild tubular atrophy)
ct2	Tubular atrophy involving up to 26-50% of the area of cortical tubules (moderate tubular atrophy)
ct3	Tubular atrophy involving in >50% of the area of cortical tubules (severe tubular atrophy)

Radboudumc